République Algérienne démocratique et populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de Saad Dahlab Blida 1
Faculté des Sciences
Département d'Informatique
Master 2 Systèmes Informatiques et Réseaux

Evolution du profil utilisateur dans un système de recherche d'information personnalisée

Réalisé par :
-M^r^ Abdelli Mohammed Adel
-M^r^ Akam Nour El Hadi

Promotrice :
M^me^ Zahra Fatima Zohra

Propose Par :
Boulkrinat Nour El Houda

Devant le jury :
-M^me^ Oukid Lamia
-M^me^ Nasri

Années universitaire : 2018-2019

Résumé

Un système de recherche d'information personnalisée a pour objectif l'amélioration des résultats retournés à l'utilisateur lors d'une session de recherche dans une grande base de documents en fonction de ses centres d'intérêt, afin de mieux répondre à ses besoins spécifiques exprimés par des requêtes.

Les intérêts de l'utilisateur définissant son profil pourraient changer avec le temps, ce qui génère le problème de l'évolution du profil utilisateur. Cette évolution du profil consiste à adapter son contenu à la variation du besoin en information. Le présent travail s'intéresse à la modélisation du profil utilisateur, et à l'enrichissement de ses centres d'intérêt à travers son activité représentée par ses requêtes émises au fil du temps, en se basant sur une des méthodes de Data Mining, plus précisément Le réseau de neurones qui a pour but d'apporter des résultats intéressants dans la classification et la génération des nouveaux centres d'intérêt.

Mots-clés : *RI, SRIP, Profil utilisateur, Requête utilisateur, Enrichissement, Réseau de neurones, Fouille de données, Application web, Django, SQLite*

Abstract

A personalized information search system aims to improve the results returned to the user during a search session in a large database of documents according to his interests, in order to better meet his specific needs expressed by queries.

The interests of the user defining this profile could change over time, which generates the problem of the evolution of the user profile. This evolution of the profile consists in adapting its content to the variation in information needs. The present work is interested in the modeling of the user profile, and the enrichment of its interests through its activity represented by its requests issued over time, based on Data Mining more precisely the neural network that aims to provide interesting results in the classification and generation of new interest centers.

Keywords: *Information retrieval, Personalized information retrieval Systems, User queries, User profile, Enrichment, Neural Network, Data mining, Web application, Django, SQLite.*

ملخص

يهدف نظام البحث عن المعلومات الشخصية إلى تحسين النتائج التي يتم إرجاعها إلى المستخدم أثناء جلسة بحث في قاعدة بيانات كبيرة تستند إلى اهتماماته، من أجل تلبية احتياجاته بشكل أفضل. طلبات محددة

قد تتغير اهتمامات المستخدم الذي يحدد ملفه الشخصي مع مرور الوقت، مما يولد مشكلة تغيير ملف تعريف المستخدم. يتكون هذا التطور للملف الشخصي من تكييف محتواه مع اختلاف الحاجة إلى المعلومات. يتناول هذا العمل الحالي نمذجة ملف تعريف المستخدم، وإثراء مراكز اهتمامه من خلال نشاطه المتمثل في استفساراته الصادرة مع مرور الوقت، بناءً على إحدى طرق استخراج البيانات، وأكثر على وجه التحديد الشبكة العصبية التي تهدف إلى تحقيق نتائج مثيرة للاهتمام في تصنيف وتوليد مجالات جديدة من الاهتمام.

***كلمات البحث**: RI، SRIP، ملف تعريف المستخدم، طلب المستخدم، التخصيب، الشبكة العصبية، استخراج البيانات، تطبيق الويب، Django، SQLite .*

Remerciement

En préambule à ce mémoire nous remercions Dieu qui nous aide et nous donne la patience et le courage, ainsi que l'audace pour dépasser toutes les difficultés.

Nous souhaitons adresser nos remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de cette formidable année universitaire.

Nous tenons à remercier notre promoteur Madame ZAHRA Fatma, et notre encadreur Madame BOULKRINAT Nour El Houda pour leurs engagements, leurs conseils, leurs orientations, leurs disponibilités, leurs aides dans le cheminement de ce mémoire et leurs œil critique nous a été très précieux pour structurer le travail et améliorer sa qualité.

A nos juges,

Madame OUKID Lamia, Vous nous faites l'honneur de présider ce jury, acceptez pour cela nos plus sincères remerciements.

Madame NASRI, Tous nos remerciements pour votre participation à ce jury.

A tous les Enseignants de l'USDB, Nous vous remercions pour tous vos enseignements qui nous ont permis de réaliser ce projet et ce, tout au long des c précédentes années.

Dédicace

Akam Nour El Hadi :

Je dédie ce mémoire à

Mes parents pour leurs éducations, leurs soutiens, tous les sacrifices consentis pour m'aider à avancer dans la vie depuis mon enfance, et leurs précieux conseils, Dieu faire que ce travail porte vos fruits et faire en sorte que jamais ne je vous déçoive.

Tous les membres de ma famille qui m'ont si lentement soutenu et cru en moi.

Mon binôme durant cette épreuve ABDELLI Mohamed Adel, Pour ta sincérité. Je te souhaite une vie pleine de joie et une carrière de succès et réussite.

Et à tous les gens présents dans mon cœur sans exception.

Abdelli Mohammed Adel :

Je dédie ce mémoire à

Mes très chers parents, qu'ils m'ont doté d'une éducation digne, leur soutien et sacrifices depuis ma naissance ont fait de moi ce que je suis aujourd'hui.

Mon oncle Saad ainsi toute sa petite famille, d'avoir toujours répondu présent. Et toute ma famille de m'avoir encouragé et soutenus durent le long de mon cycle.

Et particulièrement le précieux binôme AKAM Nour El Hadi pour sa confiance et son support inestimable. Je te souhaite tout le bonheur et la réussite dans ta vie.

Et à tous les gens présents dans mon cœur sans exception.

Sommaire

Liste des Figures

Liste des Tableaux

Introduction générale

1. Motivation et problématique

La recherche d'information est un domaine lié à la science de l'information, depuis l'apparition des premiers ordinateurs des milliers d'informations sont stockées chaque jour ; et ceux dans divers domaines de connaissances, d'où la nécessité de développer des outils performants, qui permettraient une meilleure exploitation de ces informations, tout en récoltant les données souhaitées, et éliminant celles inutiles, afin de faciliter l'accès à l'information pertinente, pour un utilisateur ayant un besoin en information.

Plusieurs travaux et recherches dans le domaine de la recherche d'informations ont été fournis donnant résultat a beaucoup d'approche qui a permet l'exploitation de celle-ci offrant des techniques et des outils afin de résoudre le problème de pertinence des résultats face au besoin en information qui sont englobé dans : les systèmes de recherche d'informations (SRI) ; ces approches qu'on l'on verra dans le premier chapitre.

Les systèmes de recherche d'information (SRI) ont pour fonction de répondre au besoin en information de l'utilisateur, qui est souvent formulé en langage naturel par une requête décrite par un ensemble de mots clés, les dernières avancées du web a remis la RI face à de nouveaux défits d'accès à l'information, à savoir retrouver une information pertinente dans un espace de taille considérable et qui répond au besoin en information spécifique de l'utilisateur.

Les premières techniques développées dans la RI étais trop statique, elles se contentent de renvoyer une liste de résultats pour une même requête semblable à tous les utilisateurs sans prendre en compte divers facteurs, ce qui a mène à l'amélioration du processus de recherche en développent des approches, qui prennent en compte la situation de l'utilisateur ce qui a donner naissance à la RI adaptative.

L'objectif de la RI adaptative consiste en la manière de modéliser l'utilisateur et l'intégration du profil résultant dans le processus d'accès à l'information. Cependant, la personnalisation de l'information engendre le problème de l'évolution du profil utilisateur au cours du temps. Dans la plupart des systèmes d'accès personnalisé, l'évolution du profil est exprimée par l'ajout de nouvelles informations**,** elle consiste à adapter le contenu du profil aux variations des besoins utilisateurs en information exprimés dans notre cas par ses requêtes, donc le changement de ses centres d'intérêts au fil du temps.

Les travaux actuels qui s'intéressent à la mise à jour du profil utilisateur, reposent sur la détection implicite des préférences et des centres d'intérêts à travers les comportements observables collectés par le système lors de l'interaction de l'utilisateur avec l'environnement (session de travail, consultation d'un document clé…etc.) et les commentaires (tags) sur les résultats retournés par le système suite aux différentes requêtes soumises.

2. Objectif

Nous nous intéressons dans ce travail, aux requêtes exprimées par l'utilisateur pour mettre à jour son profil, cette démarche est peu exploitée dans la littérature. Les techniques de Data Mining peuvent être des solutions adéquates à ce type de problème. Nous utilisons les réseaux de neurones, qui peuvent apporter des résultats intéressants dans la classification, dans le but d'enrichir les centres d'intérêts en fonction des requêtes soumises

A cet effet, notre de travail s'appuie sur la réalisation d'une application web offrant une interface à l'utilisateur, qui lui permettra de créer son profil et d'introduire ses requêtes. L'application sera gérée par un administrateur qui se chargera de mettre à jour ce profil à partir des requêtes soumises et observera les différents changements dans le système.

3. Organisation du mémoire

Afin d'aborder tous ces aspects, le présent mémoire s'articule autour de quatre chapitres :

Chapitre 1 : présente les fondements et évolution de la recherche d'information et s'appuie sur la notion du profil utilisateur.

Chapitre 2 : traite et explique la notion du Data Mining, ses méthodes et ses techniques.

Chapitre 3 : est dédié à la conception du système accompagnée d'une explication détaillée du mécanisme de l'approche choisie lors de l'enrichissement du profil.

Chapitre 4 : est consacré à la mise en œuvre et à la présentation des résultats obtenus.

Enfin, ce mémoire se termine avec une conclusion générale

Chapitre 1
Recherche d’information personnalisée

1. Introduction

Depuis l'apparition des premiers ordinateurs dans les années quarante la recherche d'informations suscite l'attention des chercheurs conscient par la valeur des documents et leur grande quantité qui ne cessera d'augmenter au fil du temps ainsi que la nécessite de les retrouver à travers de milliers ou millions d'autres.

En vue de ces défis, plusieurs approches et solutions ont été proposé dans le domaine de la recherche d'informations dans le but de stocker, organiser cette quantité des documents pour ensuite les retrouver d'une manière pertinente et rapide.

2. Recherche d'information (RI)

2.1. Définitions

Plusieurs définitions ont été introduites dans la littérature pour définir la recherche d'information nous citons :

« La recherche d'information est une branche de l'informatique qui s'intéresse à l'acquisition, l'organisation, le stockage, la recherche et la sélection d'information » **[Salton, 68]**

« La recherche d'information (RI) est un domaine qui a pour objectif d'acquérir, d'organiser, de stocker et de rechercher l'information » **[Baeza & al, 99].**

« La recherche d'information est une activité dont la finalité est de localiser et de délivrer des granules documentaires à un utilisateur en fonction de son besoin en informations. » **[Hernandez, 06]**

« La RI a pour thème central l'étude de modèles et systèmes d'interaction entre des utilisateurs humains et des corpus de documents numériques, en vue de la satisfaction de leurs besoins d'information ». **[Chiaramella & al, 07]**

2.2. Notion de base de la RI

Dans ces définition, nous retiendrons trois notions clés : « documents, requête, pertinence »

- **Document :** Un document est une collection ou un ensemble de données numérique enregistré sur un support de stockage, il peut être un texte, un morceau de texte, une page Web etc... C'est l'entité résultante en réponses à une requête.
- **Requête :** un utilisateur qui formule un besoin en information par le biais d'une requête afin d'identifier des documents pertinents en rapport à ce besoin.

- **Pertinence :** la pertinence est la correspondance entre un document et une requête, ou encore une mesure d'informativité du document à la requête. **[Boughanem & al, 03]**

3. Système de recherche d'information (SRI)

3.1. Définition du SRI

Un Système de Recherche d'Informations (SRI) est un système informatique qui permet de retourner à partir d'un ensemble de documents, ceux dont le contenu correspond le mieux à un besoin en informations d'un utilisateur, exprimé à l'aide d'une requête.

Un système de recherche d'information peut être amélioré, en modélisant, intégrant, exploitant le contexte. Ainsi, le contexte peut être utilisé par exemple pour améliorer la façon dont les individus formulent leurs besoins au sein du système de recherche d'information et explorent les informations trouvées **[Kumaran & Allan, 08].**

Système de recherche d'information est un ensemble de mécanisme de stockage, représentation et correspondance qui permet de retrouver parmi une collection, des documents pertinents en réponse à un besoin d'utilisateur exprimé sous forme d'un langage de requête. L'objectif d'un système de RI (SRI) est de relier le besoin en information d'un utilisateur, modélisé par une requête, avec un ensemble de documents en estimant leur pertinence par rapport à la requête. La qualité d'un système doit être mesurée en comparant les réponses du système avec les réponses idéales que l'utilisateur espère recevoir. Plus les réponses du système correspond à celles que l'utilisateur espère, mieux est le système, ainsi que la rapidité de délivrance des résultats ; Qui est assuré par une fonction de mise en correspondance du besoin de l'utilisateur et du corpus de documents. **[Salton & al, 83]**

3.2. Processus de la recherche d'information

Le processus de RI permet, à partir d'une requête, d'ordonnancer les documents est appelé "processus en U".

La figure ci-dessous fait ressortir des éléments constitutifs tels que : le document, le besoin en information, la requête et la pertinence, ainsi que deux principales fonctionnalités : l'indexation et appariement

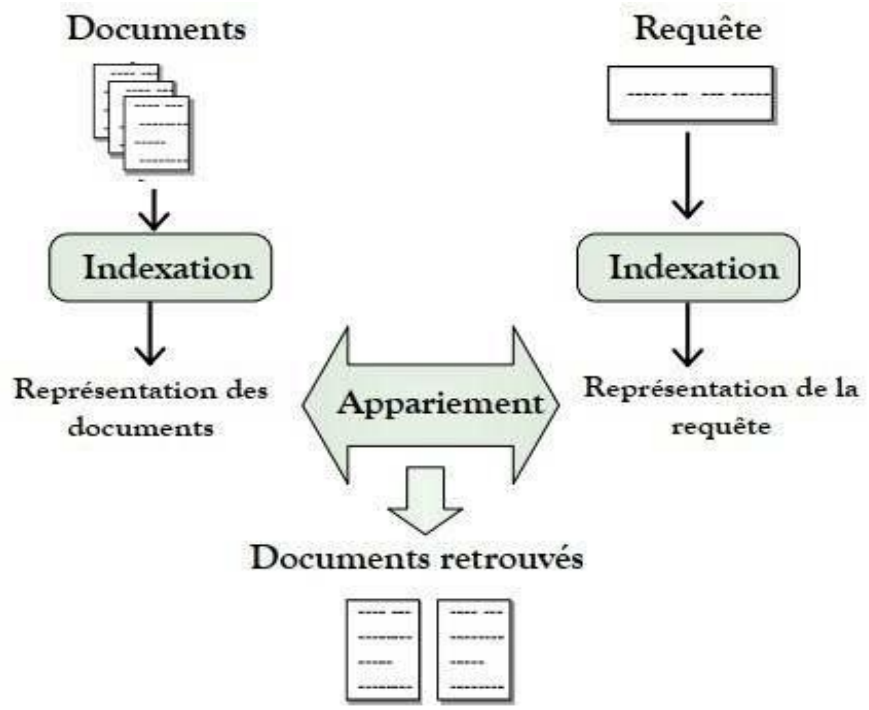

FIGURE 1.1 PROCESSUS EN U [ABBASI, 13]

3.2.1. Processus d'indexation

L'indexation consiste à extraire des documents les mots les plus discriminants encore appelés index. Cette première tâche est généralement effectuée en marge du processus de recherche car, la construction des index peut être assez longue en fonction du nombre de documents de la collection ainsi que de la taille des documents. Les index ont un caractère réducteur car tous les termes d'un document ne sont pas importants à prendre en compte pour la recherche. L'indexation peut se faire de 3 manières différentes : manuellement, de manière semi-automatique, ou de manière automatique. **[Kompaoré, 08].**

3.2.2. Processus de recherche requête-documents (Appariement)

Appariement (l'interrogation) est l'interaction d'un utilisateur final avec le SRI, une fois les documents sont représentés sous forme interne d'index. Suite à une requête utilisateur, le système calcule la pertinence de chaque document vis à vis de la requête utilisateur selon une mesure de correspondance du modèle de RI, et retourne la liste des résultats à l'utilisateur

L'interaction entre l'utilisateur et le SRI comprend : (1) la formulation d'une requête par l'utilisateur traduisant son besoin en information, (2) la représentation de la requête sous forme interne selon le langage d'indexation défini, (3) la correspondance entre la requête et les documents par exploitation de l'index et la présentation des résultats. **[Daoud, 09]**

4. Système de recherche d'information personnalisé

4.1. Définition

Un système de recherche d'information personnalisé (SRIP) est un système qui intègre l'utilisateur, en tant que structure informationnelle, tout au long de la chaîne d'accès à l'information. Le SRIP ne se limite pas seulement à modéliser les caractéristiques des utilisateurs en des profils. Il doit être capable de déduire à partir de ces profils, l'intention de l'utilisateur lorsqu'il effectue sa recherche, en d'autres termes son contexte de recherche, et de détecter l'évolution des profils de manière dynamique. Le système doit donc inclure : les préférences et les centres d'intérêts de l'utilisateur ou un groupe d'utilisateurs ainsi une procédure de mise à jour du profil qui traduit son évolution dans le temps. **[Zemirli, 08]**

4.2. Notion de Profil d'utilisateur

Le profil utilisateur dans le contexte des systèmes de personnalisation d'informations, peut être défini comme une structure qui permet de modéliser et stocker des informations relatives à l'utilisateur. Le profil utilisateur peut contenir : (les données personnelles, l'historique/ feedbacks, les annotations associées aux documents, les préférences et les intérêts) ; Les données personnelles sont relativement stables dans le temps et ne demandent pas a priori de mise à jour automatique, alors que les préférences et les intérêts tendent à changer au fil du temps. Il a pour objectif de permettre à un système de s'adapter à l'utilisateur. **[Sirinya, 17]**

4.3. Modélisation de Profil d'utilisateur

Le profil de l'utilisateur couvre des aspects larges tels que son environnement cognitif, social et professionnel qui déterminent ses intentions au cours d'une session de recherche. La plupart des travaux actuels en RI contextuelle focalisent à juste titre, sur la représentation de l'aspect lié à ces intentions qualifiées de centres d'intérêts.

Dans cette perspective, la modélisation du profil de l'utilisateur a pour objectif fondamental de représenter puis faire évoluer ses besoins en information à court et moyen terme. C'est une question qui pose la double difficulté de traduire les centres d'intérêt de l'utilisateur d'une part et faire émerger leur diversité d'autre part. Le processus de définition du profil de l'utilisateur peut être caractérisé par trois phases. La première porte sur la représentation des unités d'information représentant le profil. La deuxième phase est liée à l'instanciation de ce modèle au cours d'une activité de recherche d'information.

Enfin, la troisième phase concerne l'évolution du profil au cours du temps. Chacune de ces phases met en jeu des approches et techniques de représentation et/ou de construction. **[Hadef & al, 14]**

4.3.1. Représentation d'un profil d'utilisateur

Le profil peut prendre de différents format de présentation des centres d'intérêt de l'utilisateur, nous distinguons trois principales approches de représentation : ensembliste, sémantique et multidimensionnelle.

i. Représentation ensembliste (vectorielle)

Basé sur l'analogie au modèle vectoriel de Salton l'approche ensembliste consiste à représenter le profil de l'utilisateur par des paquets de termes pondérés ou classes de vecteurs non hiérarchisées ou hiérarchisées permettant de prendre en compte des centres d'intérêt multiples chaque intérêt est représenté par (liste de mot clé, vecteur de thème, ensemble de vecteur) **[Zemirli, 08]**

ii. Représentation sémantique

La représentation du profil met en évidence, dans ce cas, les relations sémantiques entre informations le contenant. La représentation est essentiellement basée sur l'utilisation d'ontologies **[Gauch & al, 03]** et **[Challam & al, 04]** ou des réseaux sémantiques probabilistes **[Lin & al ,05], [Wen & al, 04].**

Dans le cadre de cette approche, les centres d'intérêts de l'utilisateur sont appariés aux concepts des domaines de l'ontologie. Un profil est alors représenté en termes de concepts de l'ontologie intéressant l'utilisateur. Les ontologies de référence utilisées dans ce cadre sont basées sur la catégorisation en hiérarchie générale de Yahoo, Magellan, Lycos et ODP (Open Directory Project). **[Tamine & al ,07]**

iii. Représentation multidimensionnelle

[Kostadinov, 03] a poursuivi cette classification en proposant un ensemble de dimensions ouvertes, pouvant contenir la plupart des informations susceptibles de caractériser l'utilisateur. Dans sa représentation il distingue principalement huit dimensions décrites brièvement dans ce qui suit :

- **Les données personnelles :** elle englobe l'identité civile de l'utilisateur (nom, prénom, numéro de sécurité sociale,…) ainsi que des données démographiques (âge, genre, adresse, situation familiale, nombre d'enfants, ...)

- **Le centre d'intérêt :** exprime le domaine d'expertise de l'utilisateur. Il peut être défini par un ensemble de mots clés ou un ensemble d'expressions logiques (requêtes).
- **L'ontologie du domaine :** l'ontologie du domaine complète la définition du centre d'intérêt en explicitant la sémantique de certains termes ou de certains opérateurs employés par l'utilisateur dans son profil ou dans ses requêtes.
- **La qualité attendue :** est un des facteurs clés de la personnalisation, elle permet d'exprimer des préférences extrinsèques comme l'origine de l'information, sa précision, sa fraîcheur, sa durée de validité, le temps nécessaire pour la produire ou la crédibilité de sa source. Les attributs de cette dimension expriment la qualité attendue ou espérée par l'utilisateur.
- **La customisation :** la customisation concerne d'abord tout ce qui est lié aux modalités de présentation des résultats en fonction de la plateforme, de la nature et du volume des informations délivrées, des préférences esthétiques ou visuelles de l'utilisateur.
- **La sécurité :** la sécurité est une dimension fondamentale du profil. Elle peut concerner les données que l'on interroge ou modifié, les informations que l'on calcule, les requêtes utilisateurs elles-mêmes ou les autres dimensions du profil.

  La sécurité du processus exprime la volonté de l'utilisateur à cacher un traitement qu'il effectue.
- **Le retour de préférences :** on désigne par ces termes ce qu'on appelle communément le « feedback » de l'utilisateur .Cette dimension regroupe l'ensemble des informations collectées sur l'utilisateur.
- **Les informations diverses** : certaines applications demandent des informations spécifique ne pouvant être incluses dans aucune des dimensions précédentes comme par exemple la bande passante attribuée au gestionnaire du profil .Pour cette raison l'utilisateur a la possibilité de rajouter ce type de préférences dans la partie divers du profil et de décrire leurs utilisations. **[Zemirli, 08]**

4.3.2. Construction du profil utilisateur

Les modèles de représentations du profil d'utilisateur peuvent être simples basés sur des mots clés ou complexes basées sur des ontologies de domaines ou des hiérarchies de concepts. La construction du profil utilisateur doit se baser sur la collection et l'exploitation les différentes données et information pertinente dans le but de les représenter, la collection de ces différentes données peut se faire de manière implicite ou explicite. **[Zemirli, 08]**

- **Explicite :** elle est facile à mettre en œuvre, elle consiste à inclure l'utilisateur dans le processus en introduisant les diffèrent information et centre d'intérêt pour la construction du profil « On interroge directement l'utilisateur ou on lui demande par exemple de remplir des formulaires pour collecter les données personnelles et démographiques tels que sa date de naissance, son statut marital, son activité professionnelle et ses centres d'intérêts ». **[Zemirli, 08]**
- **Implicite :** « L'acquisition implicite ou « *feedback implicite* » consiste à collecter les données de l'utilisateur, en observant son comportement et en scrutant son activité » Elle ce fait de manière automatique en se basant sur des technique d'extraction d'information basées sur des mesures de pertinence implicite (fréquence de clics, temps de lecture, etc…). **[Zemirli, 08]**

4.3.3. Évolution d'un profil d'utilisateur

Le processus de mise à jour du profil utilisateur se fait par degrés en fonction des informations et données collecter ; L'étude de l'évolution des intérêts de l'utilisateur consiste à prendre en compte le changement de ses intérêts à travers le temps **[Crabtree & al ,98],** « La gestion de l'évolution du profil utilisateur est un processus complémentaire à la construction d'un profil utilisateur et désigne leur adaptation à la variation des centres d'intérêt des utilisateurs au cours du temps » **[Daoud, 09]**

En suivant le travail de **[Mezghani & al, 15]** le traitement de l'évolution du profil utilisateur peut se faire par deux techniques : i) l'enrichissement du profil avec de nouvelles informations de différentes techniques de détection d'intérêts, ou ii) la simplification des informations que nous considérons sans rapport et dont les valeurs diminuent au fil du temps.

I. L'enrichissement :

Est une technique qui ajoute des informations au profil utilisateur après un traitement prédéfinir.

Dans le travail de **[Zayani, 08]** la mise à jour de profil utilisateur s'applique aux attributs évolutifs, comme les intérêts et les préférences. Des mécanismes ont été utilisés pour incrémenter la valeur des intérêts selon leurs fréquences d'utilisation. **[Mezghani & al, 15]**

D'après **[Canut & al, 15]** il existe deux approches de gestion de l'évolution du profil utilisateur La première approche consiste à gérer la dynamique des intérêts de l'utilisateur après la phase d'extraction des intérêts et correspond à un processus de mise à jour du

profil utilisateur. Et la deuxième approche à qui on s'intéresse dans notre travail consiste à prendre en compte la dynamique des centres d'intérêt pendant l'étape d'extraction des intérêts. Dans ces dernières deux modèles de profil utilisateur sont utilisé, soit des modèles à court terme ou à long terme.

1) Evolution du profil utilisateur à court terme

Nous pouvons définir l'évolution du profil utilisateur à court terme par « Le profil utilisateur à court terme décrit des centres d'intérêts et des besoins utilisateurs liés aux activités et la tâche de recherche courante. » **[Daoud, 09]**

Dans la recherche d'information personnalisée, on utilise l'historique à court terme de l'utilisateur lié à une seule (la dernière) session de recherche pour extraire ses intérêts **[Bennett & al, 12]** Avec le même principe, plusieurs travaux proposent une approche utilisant un critère temporel pour mieux cerner l'évolution et la dynamique des informations étudiées. La plupart de ces travaux se basent sur l'approche « time-forgotten » qui ignore les informations trop anciennes **[Cheng & al, 08]**; **[Maloof & al, 00].**

Dans ce type d'approche, on oublie complètement les informations dépassant une date limite. Pourtant, certaines de ces informations ignorées peuvent être utiles et ne pas les prendre en compte peut entraîner une perte d'informations intéressantes. En effet, **[Tan & al, 06]** ont prouvé que l'historique de recherche à long terme est très important pour améliorer la tâche de recherche d'informations dans le cas de requêtes récurrentes. **[Canut & al, 15]**

2) Evolution du profil utilisateur à long terme

Le profil utilisateur à long terme modélise des centres d'intérêts généraux, persistants, ou récurrents. Ce profil peut-être exploitable dans le but d'améliorer la recherche pour toute requête soumise par l'utilisateur. **[Daoud ,09]**

Dans l'utilisation de modèles à long terme, toutes les informations de l'utilisateur sont conservées (et peuvent contenir éventuellement des biais), il est alors difficile de sélectionner les informations pertinentes pour représenter l'utilisateur à un instant donné. Il se peut que des intérêts anciens de l'utilisateur ne soient plus significatifs à ce jour. Cette remarque peut être retrouvée dans **[Kacem & al, 14]**; **[Li & al, 13]** qui proposent d'appliquer une fonction temporelle pour pondérer les intérêts de l'utilisateur selon leur fraicheur. Cette idée peut être retrouvée également dans le contexte de la construction du profil utilisateur à partir d'un réseau d'annotations comme dans **[Zheng & al, 11]** qui

utilise des fonctions temporelles pour pondérer des tags avant d'en extraire les intérêts de l'utilisateur. Dans ce type d'approche, toutes les informations existantes de l'utilisateur sont exploitées mais de manière plus restreinte. **[Canut & al, 15]**

II. La simplification :

La simplification du profil utilisateur élimine des informations considérées sans rapport à un utilisateur donné. Elle réduit la quantité d'informations contenue dans un profil pour faciliter le traitement. Elle permet aussi le filtrage des anciennes données qui ne reflètent plus les intérêts de l'utilisateur. Elle peut porter sur divers attributs du profil utilisateur. Les intérêts sont les attributs qui varient le plus dans un profil dans notre contexte. **[Mezghani & al, 15]**

5. Conclusion

Dans ce chapitre nous avons défini la recherche d'information RI, ses principaux concepts, ainsi le système de recherche d'information SRI. Nous avons ensuite présenté les différents processus de la RI, le processus de fonctionnement d'une RI et ses modèles, par la suite nous avons étudié le système de recherche d'information personnalisé qui inclus la présentation d'un profil d'utilisateur, modélisation et création d'un profil et ses diffèrent format.

Et pour terminer nous avons donnés un aperçu sur l'évolution d'un profil utilisateur dans un système de recherche d'information. Le chapitre suivant est consacré à une étude des différentes méthodes de Data Mining.

Apres une recherche minutieuse sur le data mining pour enrichir le profil utilisateur, et il ne y'a pas de travaux existant sur ceci dans la littérature, d'après nos connaissances notre travail est parmis le premier dans ce genre.

Chapitre 2

Fouille de données

1. Introduction

En vue de l'augmentation considérable de la quantité de données à stocker sur le web qui est accompagné par la croissance de la capacité de stockage des supports informatiques ; cela a engendré de gros volumes de données stockées chaque jour. Ces gros volumes sont des mines d'information d'où la nécessité d'avoir des outils et des techniques afin d'explorer et d'analyser d'une manière intelligente et rapide et extraire des connaissances. Le data Mining ou fouille de données donc vise à découvrir de l'information cachés en se servant de divers algorithmes issus de différentes disciplines, comme les statistiques ou l'intelligence artificielle.

2. Définition du Data Mining

Data mining est traduit littéralement par forage, fouille, exploration de donnés, c'est un ensemble d'outils qui ont pour objectifs de découvrir des connaissances sur une grande masse de données, parmi les définitions nous citons :

- Le Data Mining est un nouveau champ situé au croisement de la statistique et des technologies de l'information (bases de données, intelligence artificielle, apprentissage etc.) dont le but est de découvrir des structures dans de vastes ensembles de données. **[Hand, 00]**
- Le Data mining est un processus itératif et interactif dont l'objectif est la découverte de modèles de données valides, nouveaux, utiles et compréhensibles dans de larges bases de données **[Talbi, 15]**
- Le Data Mining, ou la fouille de données est l'ensemble des méthodes et techniques destinées à l'exploration et l'analyse de grandes bases de données informatiques, de façon automatique ou semi-automatique, en vue de détecter dans ces données des règles des associations, des tendances inconnues ou cachées, des structures particulières restituant l'essentiel de l'information utile tout en réduisant la quantité de données permettant d'étayer les prises de décisions. **[Houmadi, 07]**

3. L'Extraction des connaissances à partir des données (ECD)

L'extraction des connaissances à partir des données c'est pouvoir données du sens aux grandes quantités de données, et extraire des connaissances implicites potentiellement utiles qui sont stockées massivement d'une manière quotidienne, cette opération d'extraction nécessite une multitude de techniques, algorithmes, outils …. .

1.1. Définition :

ECD (Extraction de Connaissances à partir de Données) un processus qui permet d'identifier, dans des données, des patterns ultimement compréhensibles, valides, nouveaux et potentiellement utiles [Fayyad & al, 96] c'est l'une des définitions les plus répondues dans la communauté, mais pour pouvoir extraire des connaissances faudrait bien avoir une procédure ou démarche bien définie à suivre.

1.2. Processus d'Extraction de Connaissances à partir de Données (ECD) :

Ce processus se présente comme étant « un processus complexe, non trivial, composé de plusieurs étapes itératives, et nécessitant une interactivité permanente de la part de l'utilisateur expert. Le processus constitue une feuille de route à suivre par les praticiens lors de la planification et la réalisation des projets d'extraction de connaissances à partir de données. » [Zemmouri, 14]

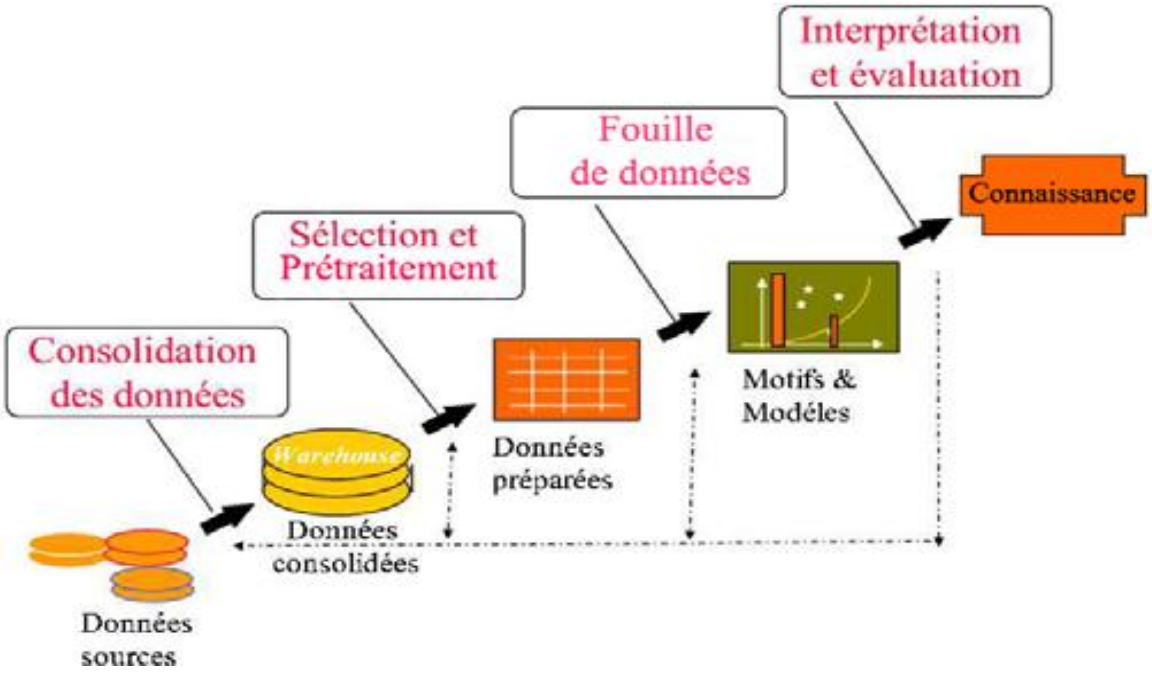

FIGURE 2.1 LE PROCESSUS DE DATA MINING [ALI LAJNEF,…,]

1.3. Etapes du processus d'ECD :

En 1996 un groupe d'analystes définit le data mining comme étant un processus composé de cinq étapes sous le standard CRISP-DM (Cross-Industry Standard Process for Data Mining) que nous allons décrire ci-dessus [El-amin, 14].

1. Définition et compréhension du problème : comprendre la signification des données et du domaine à explorer, est une étape importante, une bonne compréhension nous permettra de choisir l'algorithme adéquat à notre utilisation ainsi qu'une bonne interprétation des résultats obtenus.

2. Collecte des données : après avoir défini le problème, nous aurons une idée sur les données à collecter, des données pas toujours dans la même structure ou format, des fois nous serons amenés à collecter à partir de sources éventuellement hétérogènes (fichier, BD,..).

3.Prétraitement : les données collectées auront besoin d'être éventuellement normalisés à cause de leur incohérence, ou faire des transformations pour unifier leurs poids et permettre un traitement plus rapide, ou encore remplacer un fichier ou le supprimer carrément à cause d'anomalie causé par le système ou l'être humain.

4. Construire le modèle d'analyse : dans cette étape nous devons choisir la bonne technique pour explorer les données, pas mal existe dont les réseaux de neurones, clustering,…. généralement l'implémentation se base sur plusieurs de ces techniques pour à la fin choisir le bon résultat obtenu

5. Interprétation du modèle et établissement des conclusions (Etude des résultats) :

C'est l'étape d'interprétation du résultat et la prise de décision après visualisation des modèles qui seront plus ou moins compréhensibles à l'utilisateur pour l'aider à prendre des décisions , généralement les modèles simples sont plus compréhensibles mais moins précis, alors que ceux complexes sont plus précis mais difficiles à interpréter.

2. Méthode d'apprentissage du Data Mining

2.1. Les différentes taches du Data mining :

Le data mining selon **[Fiolet, 06]** est en réalité un ensemble d'opérations et de traitements qui mènent à la découverte de connaissances, ces diverses sont :

- **La classification :** c'est d'examiner les caractéristiques des objets et affecter les instances à leurs classes adéquates (la classe est un champ particulier a valeurs discrètes), un exemple de la tache de classification accepter ou refuser un retrait dans un distributeur. **[Lamiche, 13].**
- **L'estimation :** c'est pouvoir à partir des caractéristiques d'un objet estimer sa valeur, un champ a valeurs continues, cette estimation peut avoir comme but une classification par

exemple il suffit de définir un intervalle de valeurs pour l'attribution d'une classe particulière. [Lamiche, 13].

- **La découverte de règles d'associations :** rechercher les implications entre attributs ou ce qui les associes, cette tache a pour but d'identifier des opportunités de ventes croisées et de concevoir des groupements attractifs, plus connu sous le nom de l'analyse du panier de la ménagère, l'exemple type de cette tâche est la détermination des articles qui retrouvent sur le même ticket de supermarché. [Lamiche, 13].
- **Le clustering (segmentation) :** c'est de créer des groupes homogènes au sein d'une population, il y'a pas de classes à expliquer ou des valeurs à définir c'est au rôle de l'expert de définir la signification des groupes obtenus, cette tâche précède celles citées avant car c'est les groupes obtenu qui vont servir pour la tache de classification ou d'estimation. [Lamiche, 13].
- **La découverte de séquences :** similaire aux règles d'association avec introduction de la notion de temps. [Fiolet, 06]
- **La détection de déviation** : c'est d'identifier des valeurs exceptionnelles. [Fiolet, 06]
- **La recherche de similitudes :** identifier des instances avec des séquences communes (domaines bio-informatique). [Fiolet, 06]

2.2. Les méthodes d'apprentissage automatique :

Ce sont des techniques qui consistent à programmer une machine pour qu'elle apprenne à effectuer différentes tâches, on trouve deux genres de méthodes :

1) **Les méthodes descriptives :** qui ont pour but d'aider à comprendre les phénomènes existant, dans un ensemble de données non étiquetées [Alaoui Ismaili, 16], son objectif est donc de trouver les relations entre la caractérisation et la signification et permettant d'augmenter les connaissances du domaine [Azé, 03]. Parmi ces méthodes la segmentation (voir les différentes taches du data mining) et la description qui se concentre sur la recherche de patterns (modèles, schémas ou règles) décrivant les données interprétables par l'utilisateur [Zemmouri, 13].
2) **Les méthodes prédictives :** c'est pouvoir prévoir et expliquer à partir d'un ensemble de données étiquetées un phénomène. Ainsi chaque individu est décrit par des caractéristiques et appartenant à une classe [Alaoui Ismaili, 16], l'objectif de ces méthodes est donc de trouver une relation entre les caractéristiques permettant de prévoir et/ou expliquer le comportement de la classe, elle se décompose en deux étapes apprentissage d'un modèle sur les données de la base d'apprentissage et validation du modèle sur la partie de données dites base de validation [Azé, 03].

Parmi ces méthodes la classification (voir les différentes taches du data mining), la régression qui est similaire à la classification c'est la classe cible qui est un attribut continu [Zemmouri, 13].

3. Techniques du Data Mining

Plusieurs techniques du Data Mining sont proposées. Elles sont choisies en fonction de la nature des données et du type d'étude que l'on souhaite entreprendre, selon [Tir, 05] elle peut être sous forme d'extraction de connaissance qui est une relation très forte entre deux valeurs (implication, corrélation), ou ressemblance/similitude qui est l'ensemble des points communs à leurs descriptions. . Parmi ces méthodes on trouve les méthodes de classification comme le k-moyenne, ou par apprentissage supervisé qui offre un gain considérable de temps, réseaux de neurones, arbre de décision…

3.1. K-plus proche voisins :

L'algorithme des k plus proches voisins (k-Nearest Neighbours ou KNN) est considéré comme un algorithme paresseux du fait qu'il n'effectue aucun apprentissage contrairement aux autres algorithmes, il se construit par son modèle il se base directement sur les instances de données d'apprentissage en s'appuyant sur l'idée qu'une instance est plus proche des instances de la même classe que celles des autres [Bouaziz, 17].

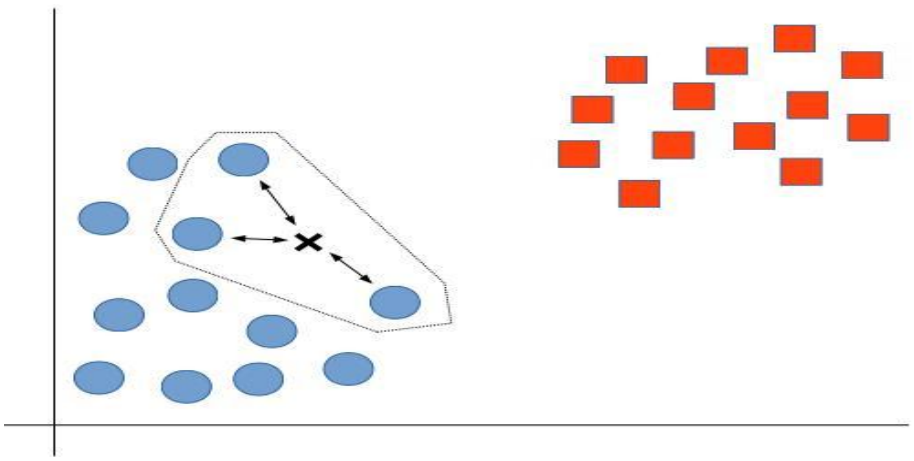

FIGURE 2.2UN EXEMPLE DE CLASSIFICATION PAR KNN AVEC K=3 [BOUAZIZ, 17].

- **Fonctionnement :**

Le principe de fonctionnement de cet algorithme est le suivant « figure 2 », on a une donnée de classe inconnue qui est comparée avec tous les données déjà stockées suivant une distance, ainsi la

nouvelle donnée est affectée à la classe majoritaire parmi ses voisins au sens de la distance [Chamroukhi, 12] ,

Cette distance offre un avantage par le fait qu'on utilisera une mesure adaptée en cas d'utilisation (la distance euclidienne est largement utilisée), le choix du K aussi à un grand effet sur les performances du classifieur [Bouaziz, 17].

3.2. Les Algorithmes génétiques AG's :

Les algorithmes génétiques sont des algorithmes d'optimisation stochastique qui travaillons sur une population basée sur les mécanismes biologiques tel que la loi de sélection, mutation … pour permettre ainsi aux ordinateurs d'imiter les êtres vivants en évoluant ; Ils sont utilisés d'après [Houmadi, 07] dans la découverte de connaissances dirigées. Ils permettent de résoudre des problèmes divers, notamment d'optimisation, d'affectation ou de prédiction.

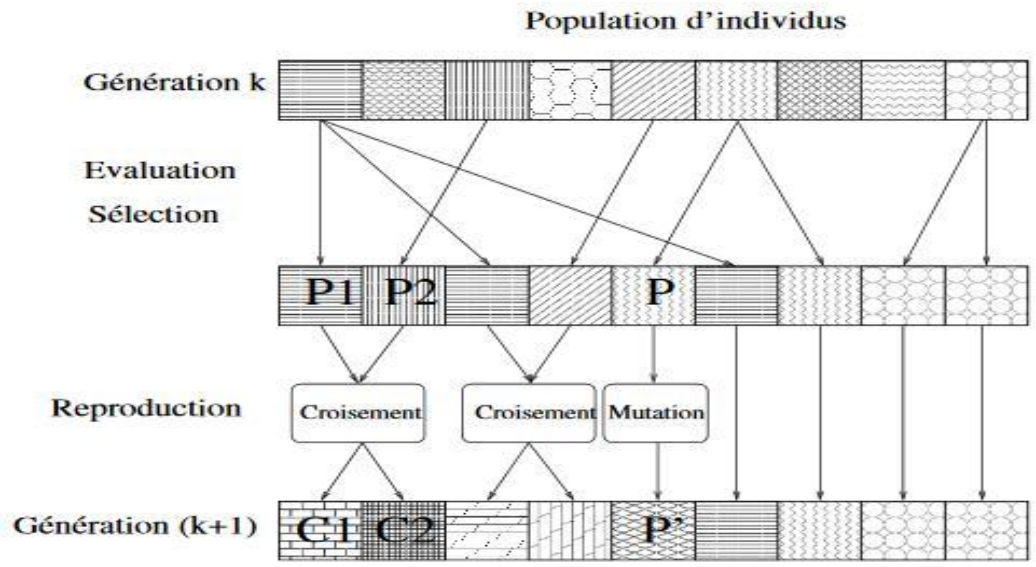

FIGURE 2.3 PRINCIPE GENERAL DES ALGORITHMES GENETIQUES [DURAND, 04]

- **Fonctionnement :**

  Les algorithmes génétiques travaillent sur une population qui contient des solutions candidates, chacune d'elle possède des caractéristiques pouvant prendre plusieurs valeurs appartenant à un alphabet non nécessairement numérique ; Le but est de trouver la meilleure combinaison de ces éléments [Khabzaoui, 06].

  D'après la figure ci-dessous « figure2 » trois opérations sont répétées pour obtenir une génération k+1, des couples parents sont sélectionnés p1/p2 en fonction de leur adéquation ; qu'ensuite une opération de croisement pour générer une couple c1/c2,

D'autres éléments sont sélectionnés en fonction de leur adéquation pour faire par la suite une opération de mutation et générer un élément p dans la génération k+1 (les individus mutés sont ensuite évalués avant insertion dans la nouvelle population) [Durand, 04].

3.3. Arbre de décision :

L'arbre de décision est une des techniques de data mining récente et efficace, elle permet d'identifier très rapidement les variables les plus discriminantes d'un jeu de données, elle offre une représentation facile à interpréter qui est sous forme d'un arbre [Santos, 15] dans un arbre « figure 4» un nœud représente un attribut spécifique et les branches constituent des conditions sur les attributs du même nœud.

Enfin les nœuds finaux sont les classe d'individu, une décision de classification est à prendre après vérifications des conditions pour l'affectation, la meilleur définition qu'on trouve est celle de **[Quinlan, 86],** «Les arbres de décision sont des architectures qui classifient les instances en entrée en les acheminant à travers des conditions posées sur les valeurs des attributs desdites instances ».

Les arbres de décision sont l'une des techniques de classification, qui peut être utilisée pour prédire les classes des nouveaux cas, ou/et extraire des connaissances potentielles à partir des données dans un but de description ou de prédiction**.** [Mededjel & al, 07]

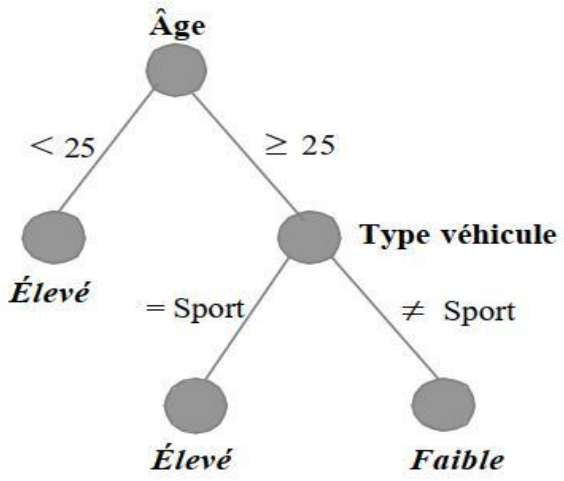

FIGURE 2.4 ARBRE DE DECISION « RISQUE ROUTIER » [MEDEDJEL, 07]

- **Fonctionnement :**

  La construction de l'arbre est faite par la sélection d'attribut qui départage les données à classifier d'une manière efficace et récursive , ainsi à la première sélection on aura l'attribut racine qui est accompagné par les conditions qui sont les branches de cet arbre pour avoir sous un nouvel attribut pour départager une autre fois les données ou une feuille qui est la classe en elle-même et ainsi de suite jusqu'à la construction de l'arbre **.[Bouaziz, 17]**

3.4. Réseau de neurones :

Les réseaux de neurones sont sujets de beaucoup de travaux, qui a comme but de simuler les transmissions nerveuses d'un être humain Larousse le définit comme étant « Cellule de base du tissu nerveux, capable de recevoir, d'analyser et de produire des informations » d'où chaque cellule neurone transmet des informations à un autre qui traite ces information et ainsi de suite pour aboutir à un résultat.

Les réseaux de neurones ont été adapté aux ordinateurs pour leur grande capacité de calcul d'où les réseaux artificiels ont vu le jour qui « sont des structures (la plu part de temps simulés par des algorithmes exécutés sur des ordinateurs d'usage général, parfois sur des machines ou même des circuits spécialisés) qui prennent leur inspiration (souvent de façon assez lointaine) dans le fonctionnement des systèmes nerveux» **[Houmadi, 07].**

1) Neurone formel :

On peut définir un neurone formel (artificiel) peut être définie étant « une unité de traitement qui reçoit des données en entrée, sous la forme d'un vecteur, et produit une sortie réelle. Cette sortie est une fonction des entrées et des poids des connexions » **[Bennaini, 14],** plusieurs neurones forment un réseau de neurone artificiel (RNN) que Touzet a défini «les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique de réseaux est évidemment un réseau » **[Touzet, 92]**

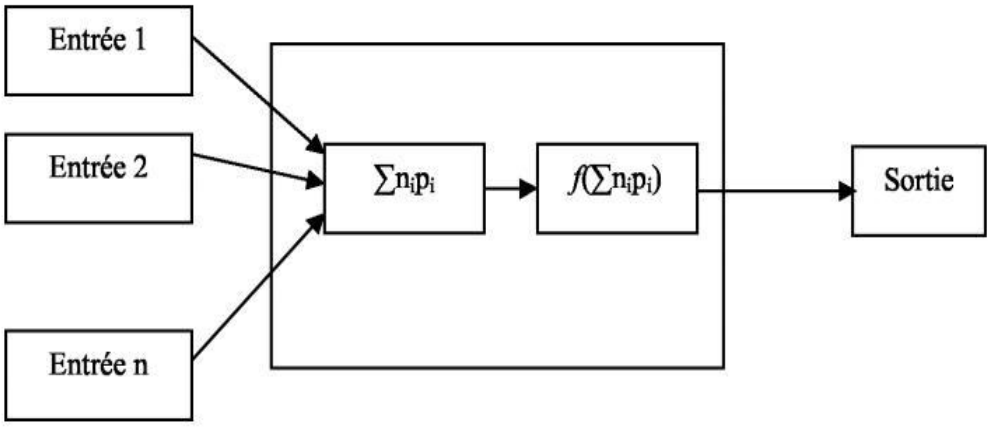

FIGURE 2.5 NŒUD D'UN RESEAU DE NEURONE [LAMICHE, 13]

A partir de cette figue nous distinguons les différentes entrées n_i ainsi que les poids p_i qui sont : « la valeur numérique du poids associé à une connexion entre deux unités reflète la force de la relation entre ces deux unités. Si cette valeur est positive, la connexion est dite excitatrice, sinon elle est dite inhibitrice » **[Bennaini, 14]**, ensuite une fonction de combinaison calcule une première valeur à partir des nœuds connectés en entrée et poids des connexions. Dans les réseaux les plus courants, les perceptrons, il s'agit de la somme pondérée $\sum_{nipi}$ des valeurs des nœuds en entrée **[Lamiche, 13],** et la fonction d'activation selon **[Houmadi, 07]** fonctionne de la manière suivante

« Si la somme des entrées est supérieure à un seuil, alors le neurone de sortie est activé ; sinon rien », mais selon Houmadi dans nos jours la plupart préfèrent utiliser des fonctions continue sur les entrées afin de permettre de traiter plus d'information dans un seul neurone.

2) Architecture d'un réseau de neurones artificiel :

On distingue dans les réseaux de neurones artificiels selon **[Kaadoud &al, 18]** deux types d'architecture qui diffère dans la façon qui sont interconnectés ainsi que le nombre de couche qui le constituent.

a. Réseau monocouches :

« Un réseau monocouche, acyclique (il ne comporte pas de boucle), la dynamique (l'activité) est déclenchée par la réception en entrée d'information. Ce réseau est dit simple car il ne se compose que de deux couches : une couche d'entrée et une couche de sortie »

b. Réseau multi couches :

« Le perceptron multicouche se compose d'une couche d'entrée, d'une couche de sortie et d'une ou plusieurs couches cachées. Si le réseau possède n couches, alors il possède n-1 matrice de poids »

Fonctionnement :

Le fonctionnement d'un réseau de neurones pour l'apprentissage et l'extraction de connaissances d'après **[Houmadi, 07]** passe par les phases suivantes :

1. La construction de la structure d'un réseau neuronal.
2. La construction d'une base pour l'apprentissage, qui est constituée de vecteurs qui représente le domaine à modéliser.
3. La phase d'apprentissage où on présente au réseau les différents vecteurs constituent la base d'apprentissage, deux modes d'apprentissages,
   a. **les réseaux supervisés :** « Le réseau apprend par présentation de pair d'entrée/sortie. Durant l'apprentissage, les valeurs de sorties désirées sont comparées à celles produites par le réseau. L'erreur résultante est utilisée pour l'ajustement des poids des connexions. » **[Kalakh, 2013]**
   b. **les réseaux non supervisés :** « aucune information sur la sortie désirée du réseau n'est disponible. Ainsi, le réseau manipule des données qui lui sont présentées en entrée et cherche à extraire quelques propriétés qui formeront les sorties du réseau. L'extraction de ces propriétés dépend de la règle d'apprentissage utilisée dans le réseau » **[Kalakh, 2013]**

4. Conclusion

Dans ce chapitre nous avons présenté la définition du Data Mining, ainsi l'extraction de connaissance, par la suite on a défini les différentes méthodes d'apprentissage.

A la fin nous avons cite quelques techniques du Data Mining qui font la classification à partir d'un apprentissage supervisé.

Chapitre 3

Approche proposée

1. Introduction

L'introduction du profil utilisateur ou plus précisément les centres d'intérêts dans le processus de recherche d'information a pour but de permettre un accès plus pertinent aux documents et pour satisfaire le besoin en informations de l'utilisateur, le problème qui intervient généralement dans ce genre de processus c'est de pouvoir maintenir à jour le profil utilisateur au fil du temps après son interaction avec le système d'une manière implicite. Pour ce faire, nous avons proposé d'utiliser une méthode de classification (les réseaux de neurones artificiels) afin d'enrichir le profil d'utilisateur dynamiquement.

2. Architecture du système

La figure ci-dessus donne un aperçu de l'architecture générale de notre système où nous distinguons les différents composants qui sont : acquisition du profil, recherche, enrichissement et statistique.

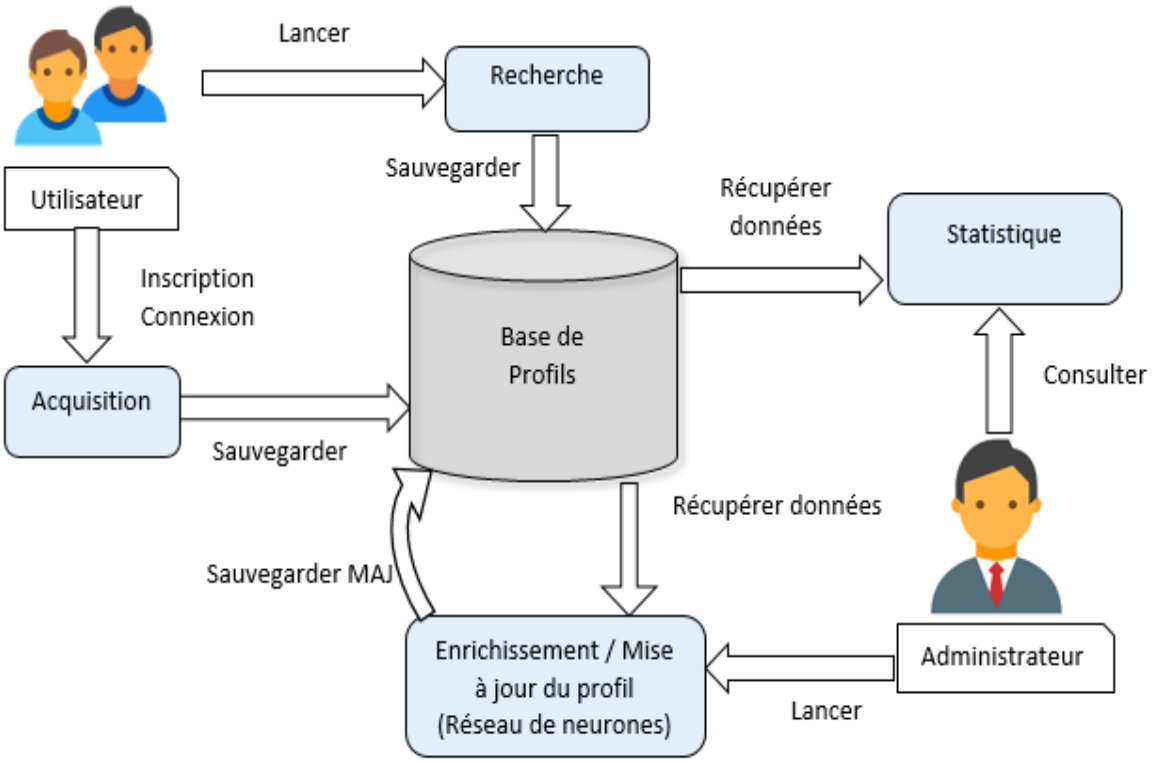

FIGURE 3.1 ARCHITECTURE DU SYSTEME

2.1. Acquisition d'un profil utilisateur

Lors de chaque nouvelle inscription l'utilisateur est invité à fournir ces informations personnelles, ces préférences et ces centres d'intérêts (optionnellement) (acquisition dites explicite), ces derniers seront utilisés pour permettre lors des sessions de recherche de retourner des résultats pertinents en fonction de la requête formulée.

Les centres d'intérêts auront après l'inscription d'un membre un degré d'importance (poids) par défaut et similaires entre eux, ensuite à l'interaction de l'utilisateur avec le système ils seront mise à jour (réajustement de poids, ajout/suppression de centres).

Les poids seront par défaut calculés la première fois selon la formule suivante :$p = \frac{1}{nbrCI}$;

Tel que : nbrCI est le nombre des centres d'intérêt d'un profil ; et $p \in \{0,1\}$

2.2. Recherche

L'utilisateur pour avoir accès aux documents voulus devra exprimer son besoin d'information sous forme d'une requête, ces requêtes seront enregistrées dans la base de données qui vont servir ensuite au processus d'enrichissement du profil utilisateur.

Pour notre travail, nous intéresserons à la mise à jour du profil de l'utilisateur en fonction de son activité dans le système.

2.3. Statistiques

L'administrateur aura la possibilité de visualiser les différents profils utilisateurs et voir leurs statistiques, ainsi que consulter leurs différentes activités.

2.4. Enrichissement

L'enrichissement c'est le cœur de notre travail il est lancé par l'administrateur, c'est un processus complexe qui va maintenir les centres d'intérêts à jour en fonction des recherches faites par l'utilisateur, elle est faite d'une manière automatique et invisible à l'utilisateur grâce à notre système, ce processus pourra en conséquent l'ajout ou la suppression des centres, ou encore l'ajustement du poids des centres d'intérêts.

Pour cette partie nous avons opté pour une méthode de data mining qui est les réseaux de neurones artificiels, que l'on verra en détail dans la section suivante :

3. Mise à jour des centres d'intérêts

Un des problèmes qu'on rencontre souvent est la mise à jour automatique des centres d'intérêts d'un utilisateur et nous utiliserons les réseaux de neurones en modélisant et implémentant la solution suivante, dans le but de mettre à jour les profils utilisateurs et d'avoir à tout moment un vecteur de centre d'intérêts adéquats et conforme aux nombreuses requêtes émises par l'utilisateur :

Un réseau de neurones dans notre cas prend comme entrées les différentes requêtes de l'utilisateur et son vecteur de centre d'intérêt initial et classe chaque couple : (centre d'intérêt depuis le requeté, vecteur initial de centres d'intérêts) dans l'une des classes suivantes : Ajout, Ajustement.

Ce réseau s'adaptera à l'utilisateur et fournira une action à entreprendre après chaque requête pour permettre la mise à jour de son vecteur.

La fonction sera faite ainsi :

En entrées : un ensemble de requêtes (qu'on va par la suite extraire les centres ainsi que leur occurrences dans les requêtes), le vecteur de centres d'intérêts de l'utilisateur.
-On fait passer chaque requête avec le vecteur dans le réseau de neurones pour qu'il puisse classer la situation dans l'une des classes : Ajout, Ajustement.

-Une fois classée l'une des fonctions suivantes se lancera : Ajout (centre d'intérêt, vecteur) ou bien Ajustement (centre d'intérêt, vecteur).Selon le résultat du classement.
-Finalement une fois toutes les requêtes traitées on lance la fonction Supprimer (vecteur de centres d'intérêt).

En Sortie : un vecteur de centres d'intérêts mise à jour.

Et pour que le réseau soit opérationnel il faut passer par les étapes suivantes :

- Création du modèle.
- Choix des hyperparamètres (nombre de couches, type du modèle...).
- Choix de la fonction de coût.
- Entrainement du modèle.

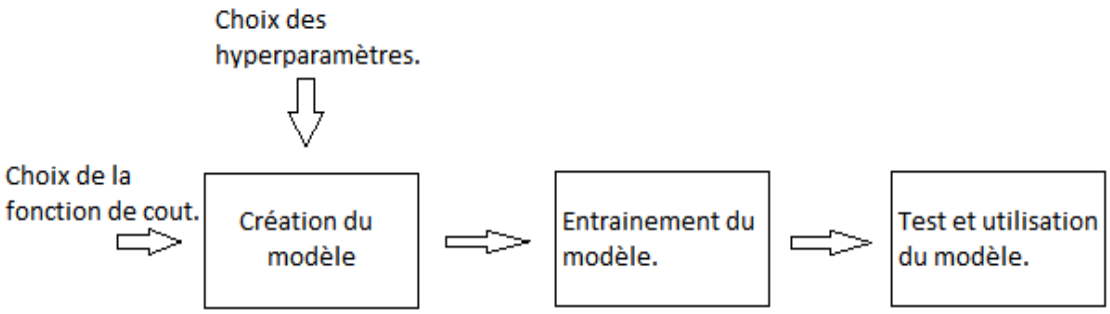

FIGURE 3.2 RESUME DE LA FONCTION DE RESEAU DE NEURONES

- Tests et utilisation.

La procédure totale se résume en ce schéma :

Implémentation du réseau de neurones dans le projet :

Pour implémenter un réseau de neurones il nous faut les éléments suivants :

- Les données d'apprentissage sous format (.csv).
- Un model qui englobe les traitements du réseau de neurones.

Notre implémentation est expliquée comme suit :

3.1. Création du model :

Pour la création de notre modèle ou boite noire, qui va accueillir les neurones et les organiser sous forme de couches de neurones reliés entre eux, par des relations logiques là ou chaque couche traite les données entrantes de la couche qui la précède.

- Récupération du contenu du document XML pour le parcourir et en extraire les données.
- Itération sur les éléments du fichier XML pour la récupération des données des requêtes de l'utilisateur « l'Id, le poids, le contenu de la requête et le centre de l'utilisateur ».
- Exportation des données sous format « csv » sous le nom de cleaned.csv

Après la récupération des données d'apprentissage sous un format exploitable faut passer à l'activation de l'environnement et la récupération des données et on fait un casting sur les données qui ont un format TensorFlow pour en changer le type.

Apres cette étape nous passerons à la création concrète du modèle en utilisant le module « keras » de TensorFlow, on va créer un modèle séquentiel ce qui veux dire que le passage par couche du réseau de neurones sera séquentiel l'un après l'autre.

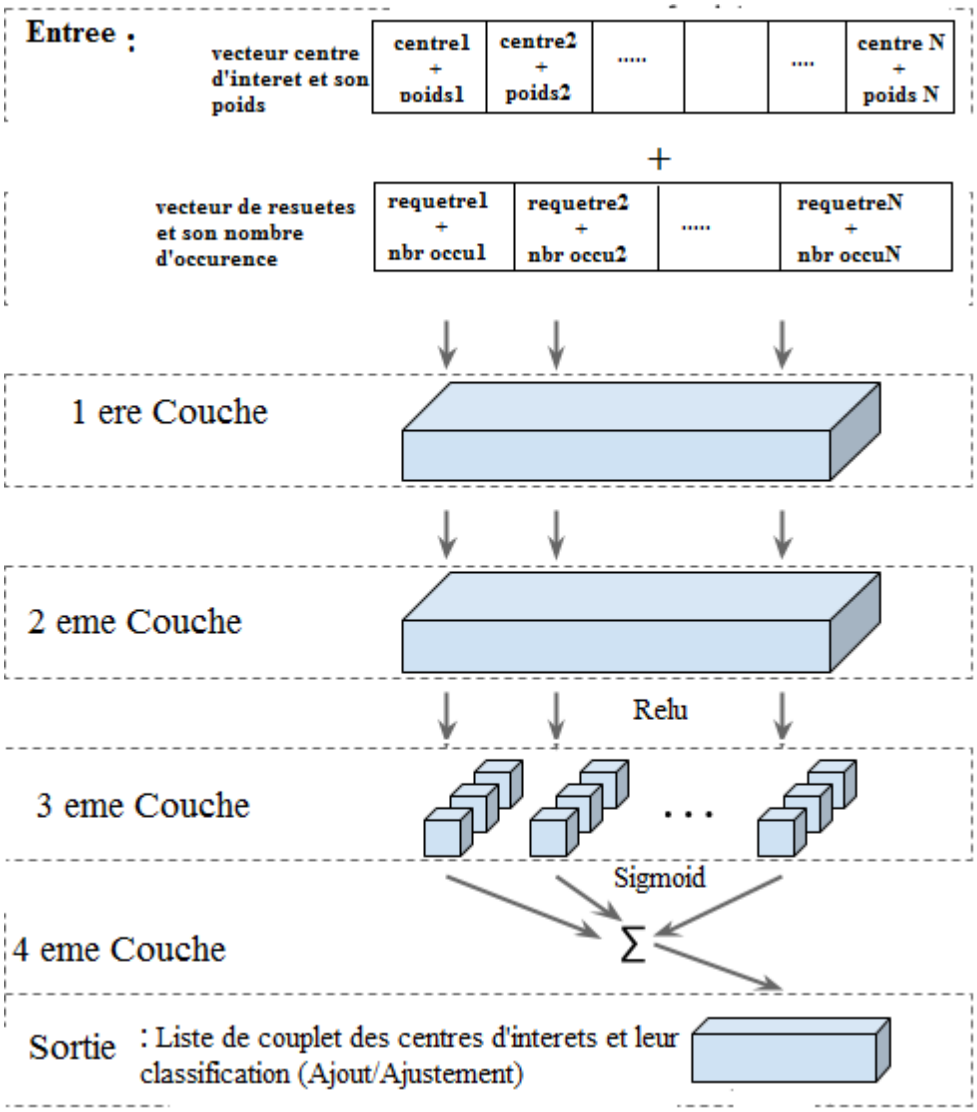

FIGURE 3.3 3 SCHEMATISATION DE LA CREATION DU MODEL

On le crée en passant comme paramètres une liste de couches :

- **Première couche** : spécification de la forme des données en entrée du réseau (taille du vocabulaire, dimension d'inclusion et la taille de l'entrée.

- **Deuxième couche** : qui unifie les données de dimension 1 car on n'a pas des données matricielles.
- **Troisième et quatrième couches** : une couche de réseau de neurones densément connecté qui applique en sorti une fonction d'activation sur les données.

Ensuite nous passerons à la Compilation du modèle : qui consiste à configurer le processus d'apprentissage qui comporte le choix de l'optimiseur (optimisation stochastique « Adam »), la fonction qui calcule la perte d'information (loss_function) et les métriques du modèle (pour la classification qui est notre besoin utilise « accuracy » comme métriques) :

- Métrique :

  En règle générale accuracy est utilisé, Comme pour la fonction de perte, nous définissons également une métrique pour le modèle. De manière simple, les métriques peuvent être comprises comme la fonction utilisée pour évaluer les performances du modèle sur un jeu de données, également appelé jeu de données de validation. La seule différence entre les métriques et la fonction de perte est que les résultats des métriques ne sont pas utilisés pour l'optimisation ou l'ajustement du modèle. Ils ne sont utilisés que pour valider les résultats du test.
- Optimiser **:**

  La fonction d'optimisation est un algorithme mathématique qui utilise des dérivées, Il sert à comprendre les changements qui se sont déroulé dans le réseau de neurones et suivre l'évolution.

  Les changements seront faites grâce à la fonction de perte qui va à chaque itération modifier les poids de connexion entres les neurones, soit en diminuant ou en augmentant le poids entre les neurones pour un meilleur résultat, notre choix c'est porter vers Adam qui est de loin l'un des optimisateurs le plus utilisé.

  Le modèle doit savoir à quelle forme d'entrée il doit s'attendre. Pour cette raison, la première couche d'un modèle séquentiel doit recevoir des informations sur la forme en entrée.

3.2. Entrainement du modèle :

Après la configuration du model pour la classification on l'entraine sur les données en utilisant la méthode « fit ».

Voici un schéma qui explique le déroulement de la phase d'entrainement sur le réseau de neurone

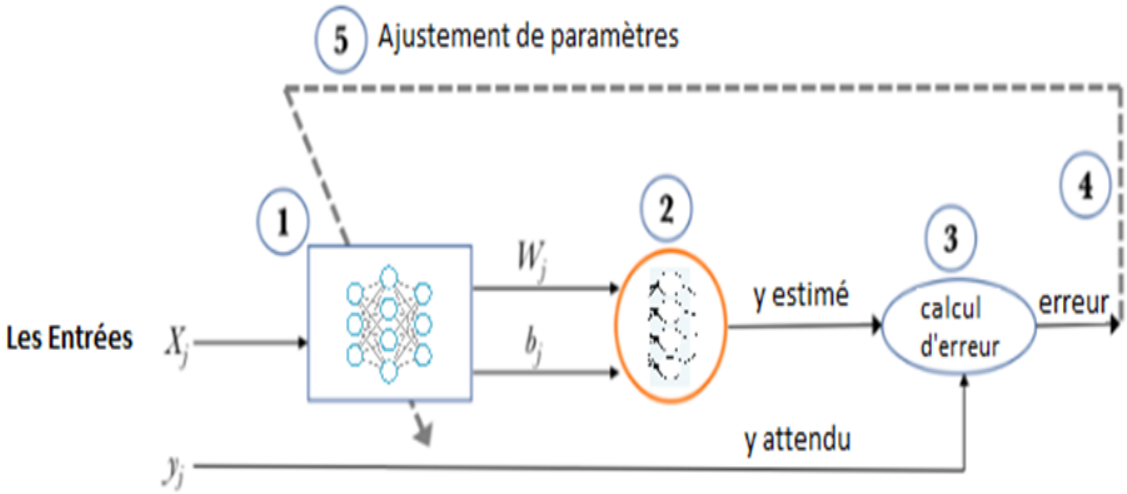

FIGURE 3.4 ENTRAINEMENT DU MODELE

1) La première étape est la phase de forward propagation

Cette étape est faite durant l'apprentissage sur les données en parcourant ou passant par tous le réseau de neurones (toutes les couches du réseau de neurones) donc chaque neurone dans une couche traite et transforme les donnes en entrée et passe l'information à la couche suivante jusqu'à la dernière couche qui prédit la classification.

2) La deuxième phase loss function

On utilise une fonction qui estime la perte ou bien l'erreur pour mesurer la précision de la classification, Par conséquent, au fur et à mesure de la formation du modèle, les poids des interconnexions des neurones seront ajustés progressivement jusqu'à l'obtention de bonnes prédictions (une erreur presque nulle), cette phase d'ajustement s'appelle backward propagation.

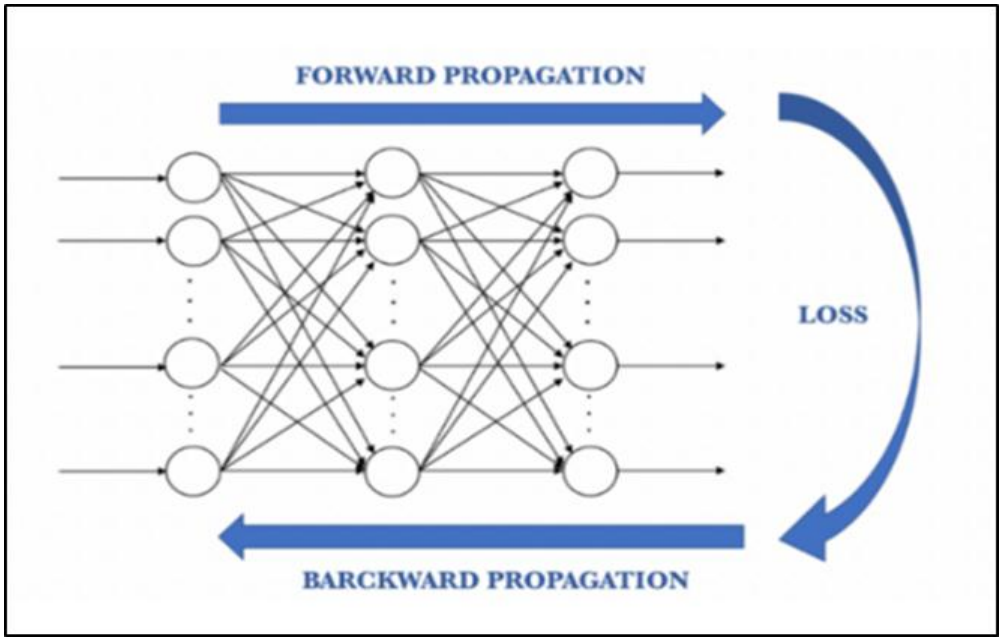

FIGURE 3.5 RESUMER DES ETAPES DE LA FONCTION RESEAU DE NEURONES

3.3. Tests et utilisation

Voici un aperçu sur le travail du réseau de neurones et l'amélioration de l'information qu'il produit.

- Dégradation du taux de perte et évolution de la qualité de l'information :
- Pendant le déroulement de la fonction de réseau de neurones le résultat obtenu s'améliore et le taux de perte diminue après chaque itération ce qui rend la classification plus crédible.

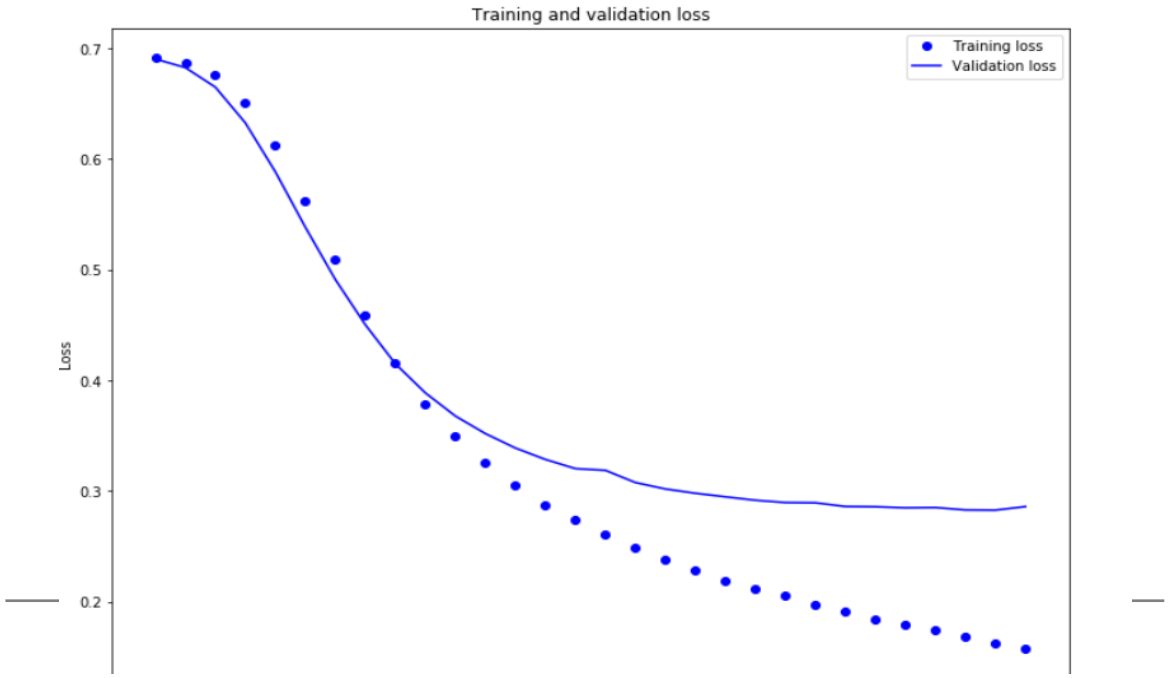

FIGURE 3.6 TAUX DE PERTE QUI DIMINUE APRES CHAQUE ITERATION (EPOCH)

- Une fonction de cout (costFunction) calcule la différence entre les prédictions du modèle et les données d'entrainement, ce qui nous permet de savoir si l'entrainement ce passe bien ou non, il faut que la valeur de cette fonction diminue après chaque itération, car après chaque itération l'algorithme essaye d'ajuster le modèle en modifiant ces paramètres d'où la diminution de l'écart entre les prédictions du modèle et les données de l'entrainement.

Le graphe de cette fonction est une courbe quadrique donc elle accepte un minimum global qui sera atteint durant l'apprentissage, pour l'entrainement du modèle on commence par des paramètres aléatoires avec un cout X qui diminue au fur et à mesure qu'on ajuste les paramètres.

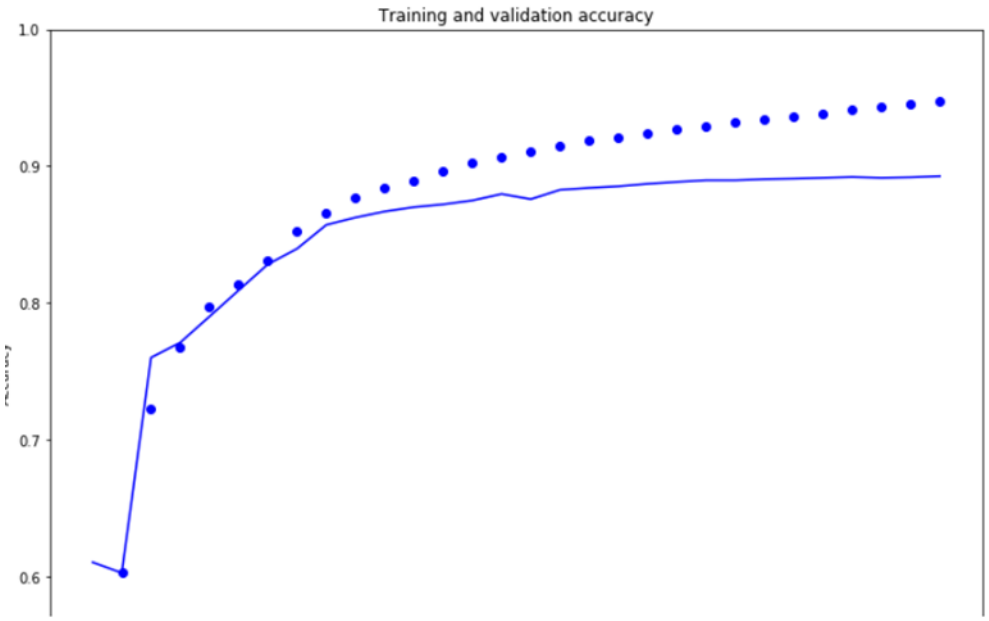

FIGURE 3.7 EVOLUTION DE LA QUALITE DU RESULTAT (ACCURACY)

- L'accuracy est la précision du modèle, Elle dépend de la complexité du modèle (spécifique ou généraliste), de son adéquation et de son adaptation au sujet à traiter. Pour cela on devise les données d'apprentissage en deux training set et test set (les données d'apprentissage et les données de test), et on teste le modèle sur les données de test à chaque itération, si notre algorithme est bien conçu on doit remarquer une amélioration continue de la précision jusqu'à atteindre une valeur seuil (0.9 = 90% ici).

- Distribution et regroupement des éléments de la même classe.

On peut selon le graphe suivant distinguer les éléments (centres d'intérêts dans la requête qui ont la même classe) par exemple (python, SQL, machine Learning …).

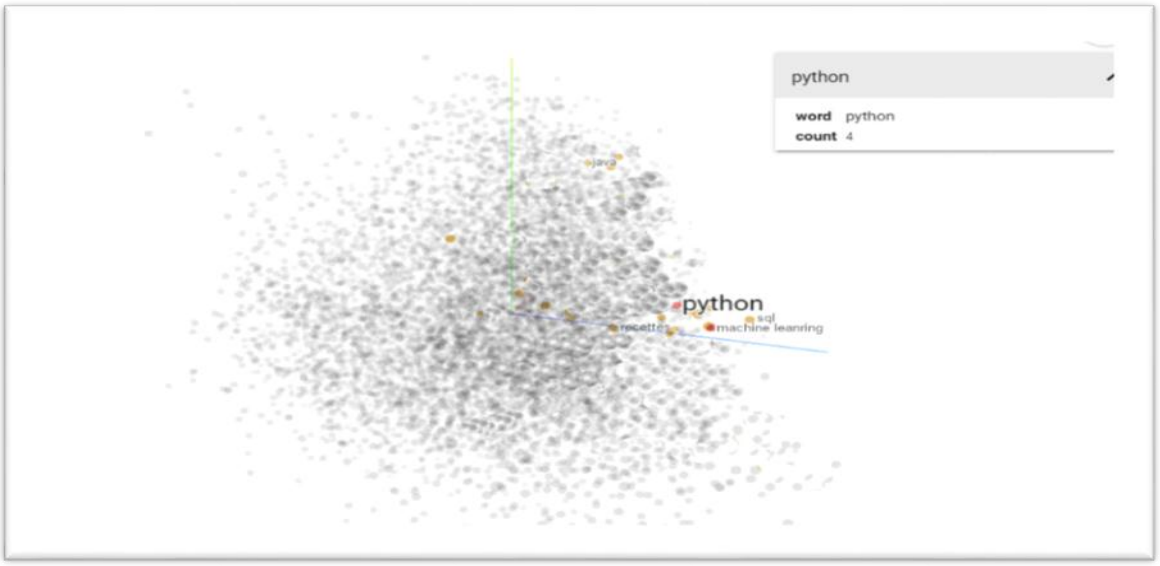

FIGURE 3.8 DISPERSION ET REGROUPEMENT DES TERMES

1	java	0.4	"oracle"
1	"java"	0.4	"oracle add"
1	"java"	0.6	"java ajust"
1	"java"	0.4761	"machine learning add"
1	"java"	0.5239	"python add"
2	"recettes	0.5	"recettes add"
5	"data min	1	"data mining ajust"
3	"python"	0.5239	"python" add
3	"python"	0.4761	"machine learning adjust"
4	"security"	1	"security add"

FIGURE 3.9 FICHIER APRES LE LANCEMENT DE LA FONCTION DE RESEAU DE NEURONES

Après la classification des centres d'intérêts dans une des classe suivants ajouts ou ajustement une fonction doit être appelée pour le calcul du poids en fonction de la classification

Ajout :

si un centre d'intérêt est juge pertinent et qui n'existe pas dans les centres du profil utilisateur son poids serra calcule selon la formule suivante :

$$Ajout = \frac{\frac{nbr_{occurence} * 0.7}{nbr_{centre} + 1}}{1 + \frac{nbr_{occurence} * 0.7}{nbr_{centre} + 1}}$$

ÉQUATION 1 AJOUT D'UN CENTER

Ajustement :

si un centre d'intérêt est juge pertinent et qui existe déjà dans les centres du profil, son poids sera calculer selon la formule suivante :

$$Ajustement = \frac{ancien\ poid + \frac{nbr\_occurences}{nbr\_centres}}{1 + \frac{nbr\_occurences}{nbr\_centres}}$$

ÉQUATION 2 AJUSTEMENT D'UN CENTRE

Apres le calcul de poids faudra recalculer les poids des autres centres d'intérêt pour respecter la contrainte de l'addition de tous les poids des centres d'intérêts doit être égale à 1 ; Ainsi ils seront calculé selon cette formule :

- Si Ajustement :

$$Sum_{div} = 1 + \frac{nbr\_occurence\ du\ centre}{nbr\_centres}$$

ÉQUATION 3 CALCUL DES POIDS SI AJUSTEMENT

- Si Ajout :

$$Sum_{div} = 1 + poids\ du\ centre\ ajouté$$

ÉQUATION 4 CALCUL DES POIDS SI AJOUT

- Au final : calcul du reste des poids des centres

$$Poids\ des\ autres\ centres = \frac{poids\ du\ centre}{sum\_div}$$

ÉQUATION 5 CALCUL DU RESTE DES POIDS DES CENTRES

Ainsi tous les centres d'intérêts avec un poids inferieur a ce seuil la se verront supprimer du profil utilisateur.

Enfin pour garder que les centre d'intérêt pertinents une fonction de suppression est invoquer à la fin ou elle doit calculer un seuil par utilisateur selon la formule suivantes

$$Seuil = \frac{1}{nbr_{requetes} * 10}$$

ÉQUATION 6 CALCUL DU SEUIL

3.4. Diagramme de classe

Le diagramme de classes représente, de manière statique, les classes qui composent le système, ainsi que les relations existant entre elles, il a pour but de décrire la structure statique interne précise de chacune des classes (attributs et méthodes), ainsi les relations entre les classes mise en œuvre.

La construction du diagramme de classe qui est unique, se fait en partie à l'aide des informations issues des différents diagrammes de séquence, il permet d'obtenir le squelette du code par génération automatique de code ; il s'agit donc de la dernière étape d'analyse juste avant le codage proprement dit.

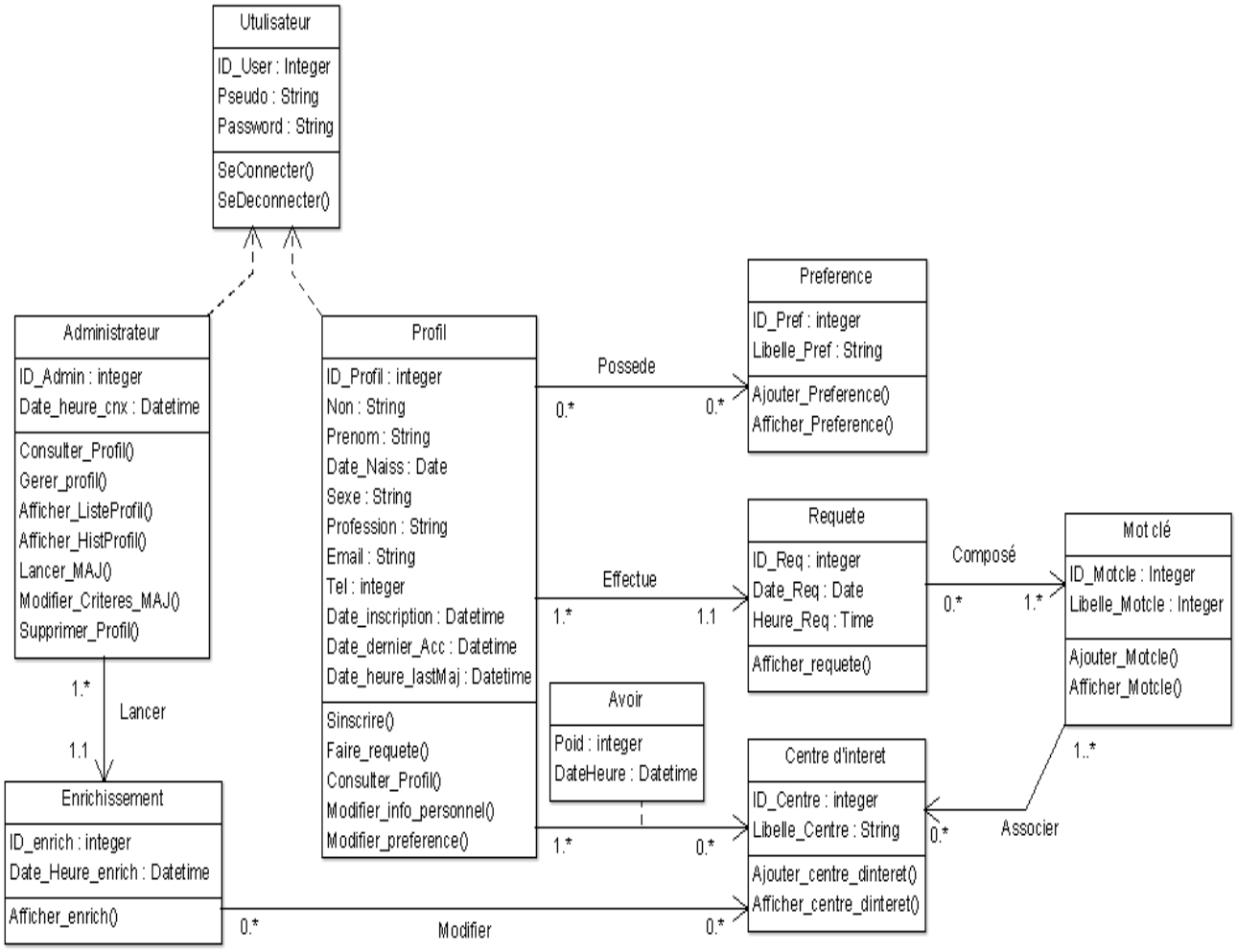

FIGURE 3.10 DIAGRAMME DE CLASSES

i). Description des classes du système

Utilisateur	Cette classe représente une personne ayant un compte dans le système.
Administrateur	Cette classe identifie les gérants du système
Profil	Cette classe identifie les utilisateurs du système
Préférence	Cette classe représente les préférences des utilisateurs (langue, format, ...etc.)
Requête	Cette classe représente les requêtes effectuées par les utilisateurs
Mot clé	Cette classe représente les mots-clés introduits dans les requêtes
Centre d'intérêt	Cette classe représente les centres d'intérêt des utilisateurs
Enrichissement	Cette classe représente les mise à jours lancé par les gérants

ii). Description des relations entre les classes

TABLEAU 3.1 ME DE CLASSERELATIONS DU DIAGRAM ESCRIPTION DESD

Association	Type	Classe A	Classe B	Description
Lancer	1.*	Administrateur	Enrichissement	Un administrateur lance un ou plusieurs enrichissements
	1.1	Enrichissement	Administrateur	Un enrichissement est lancé par un et un seul administrateur
Modifier	0.*	Enrichissement	Centre d'intérêt	Un enrichissement modifie plusieurs ou aucun centres d'intérêt
	0.*	Centre d'intérêt	Enrichissement	Le centre d'intérêt est modifié durant plusieurs ou aucun enrichissements
Possède	0.*	Profil	Préférences	Un profil contient plusieurs ou aucune préférences
	0.*	Préférences	Profil	Une préférence est associe à plusieurs ou aucun profil
Effectue	1.*	Profil	Requête	Un profil effectue au moins une requête

	1.1	Requête	Profil	Une requête est effectuée par un seul et unique profil
Avoir	1.*	Profil	Centre d'intérêt	Un profil doit avoir au moins un centre d'intérêt
	0.*	Centre d'intérêt	Profil	Un centre d'intérêt est associé à plusieurs ou aucuns profils
Composé	0.*	Requête	Mot clé	Une requête est composé de zéro ou plusieurs mot-clé
	1.*	Mot clé	Requête	Un mot-clé se trouve au moins dans une requête
Associer	1.*	Mot clé	Centre d'intérêt	Un mot-clé est associé au moins dans un centre d'intérêt
	0.*	Centre d'intérêt	Mot clé	Un centre d'intérêt à un ou plusieurs mots clé.

4. Conclusion

A travers ce chapitre nous avons détaillé l'architecture de notre système et les composants principaux qui le constituent ainsi que les acteurs qui seront en interaction avec, ensuite nous avons passé à la phase importante de notre système qui est l'utilisation des réseaux de neurones pour le processus d'enrichissement, nous avons détaillé celle-ci.

Enfin, les étapes pour entamer l'implémentation de notre système se fera dans notre prochain chapitre.

Chapitre 4

Implémentation

Et Mise en œuvre

1. Introduction

Durant le précèdent chapitre nous avons abordé la conception de notre site à travers de nombreux diagrammes et expliquer le fonctionnement du module d'enrichissement du profil, que nous avons détaillé pour la partie des réseaux de neurones, pour ce chapitre.
Nous verrons les différents outils que nous aurons à utiliser pour la création de notre site web ainsi que sa mise en œuvre ce chapitre sera composé d'une partie ou nous parlerons de l'environnement de développement choisi et une 2eme partie ou ça sera la présentation des différentes interfaces que constitues le site.

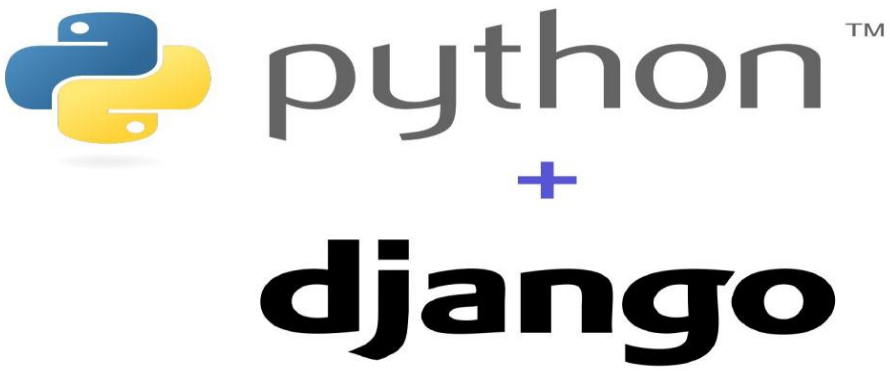

FIGURE 4.1 ENVIRONNEMENT DE DEVELOPPEMENT

2. Environnement de développement

Nous verrons à la suite de cette partie les différents outils que nous avons choisis pour la création de notre site

2.1. Langage de programmation

Pour le langage de programmation nous nous sommes orientés vers python pour plusieurs raison et avantage qu'offre par apport à notre besoin :
-langage simple syntaxiquement et facile à prendre en main
-langage de programmation interprété, du coup pas besoin d'attendre dans un projet la compilation de 10000 lignes de code celle-ci se fera au fur et à mesure
-langage fourni avec différentes bibliothèques (adapter aux machines Learning) avec possibilité d'ajouter d'autre, pour étendre ces possibilités (très grandes variétés)

2.2. Framework Django (2.0)

Django est un Framework (traduit littéralement par cadre de travail) open source destiné au web, comme tous les Framework Django permet le développement d'une manière plus rapide et rendre son code modulable et facilement réutilisable par autrui du fait de sa structuration. Django offre aussi la possibilité de génération automatique d'espace d'administration/membres, il propose aussi des outils spécialisés et très pratique pour l'interaction avec les bases de données comme ORM « object-relationnal mapping », sans oublier la grande communauté derrière Django qui minimise les difficultés pour la documentation.

2.3. Modèle MTV

Le Framework Django utilise l'architecture **MVT** (Modèle-**V**ue-**T**emplate) ; ce modèle de structuration du code est similaire au modèle connu sous le nom MVC, ainsi le modèle gère l'interaction avec la base de données,

Modifie et enregistre des informations en plus d'autre traitement sur ces données, la vue comme son nom l'indique c'est la partie que l'utilisateur verra, la différence avec le MVC réside dans le Template qui est un fichier qui contient du HTML ce fichier sera récupéré par la vue pour l'afficher cependant avant d'être envoyé au visiteur il sera analysé et exécuter par le Framework, Django a son propre moteur de Template (*Django Template Engine*).

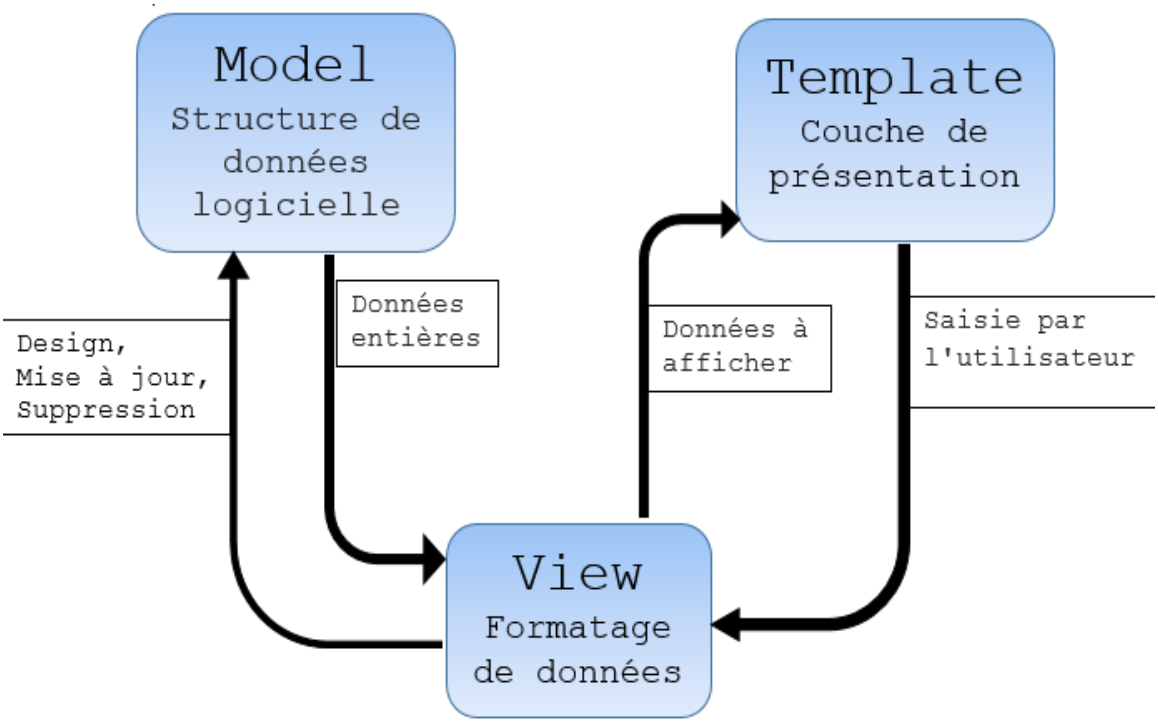

FIGURE 4.2 MODELE MTV

2.4. TensorFlow :

Est une bibliothèque de machine Learning open source développer par Google, il a été conçu pour pallier les difficultés à manipuler les technologies du Depp Learning en offrant une boite à outil qui permet de développer des architectures d'apprentissage expérimentales en outre d'entraîner et d'exécuter des réseaux de neurones pour la classification, et de résoudre des problèmes mathématiques complexes, elle offre un grand nombre de modèle et algorithme dédié au Depp Learning .

2.5. SGBD SQLite

SQLite est un SGBD que contrairement aux autres SGBD, tel que MySQL il ne se base pas sur le modèle client-serveur mais il est directement intégrée aux programmes. L'intégralité de la base de données (déclarations, tables, index et données) est stockée dans un fichier qui est indépendant de la plateforme.
Il est intégré dans les bibliothèques standards de beaucoup de langages comme PHP ou Python, il est connu aussi par son extrême légèreté (moins de 300 Ko).

3. Mise en réseau

3.1. Interface

Les différentes interfaces du site, les zones d'interaction des utilisateurs avec notre système, nous présentons en ce qui suit les différentes pages mises à disposition des membres/administrateurs et les principales fonctionnalités.

- **Accueil**

Notre site se base sur un système d'authentification, afin de définir un profil d'utilisateur unique pour chaque membre qui va servir dans l'enrichissement de ses centres d'intérêts en fonction de ses recherches,

Ainsi dans la page d'accueil dans le cas où le visiteur est non inscrit il sera invité à s'inscrire ou il sera dirigé vers la page d'inscription, dans le cas contraire il aura à entré dans l'espace adéquat ses identifiants afin d'accéder à son espace personnel pour effectuer ses recherches ou accéder à d'autres fonctionnalités.

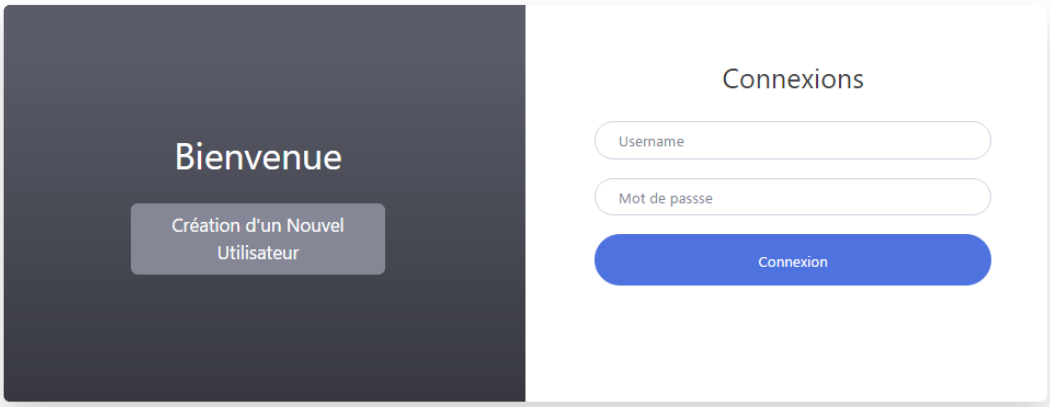

FIGURE 4.3 ACCUEIL

- **Inscription**

  lors de l'inscription d'un nouveau profil le futur membre aura à introduire ses différentes informations personnel (obligatoirement) et ses préférences, centres d'intérêts (optionnellement) cette dernière qui va servir dans le processus de mise à jour.

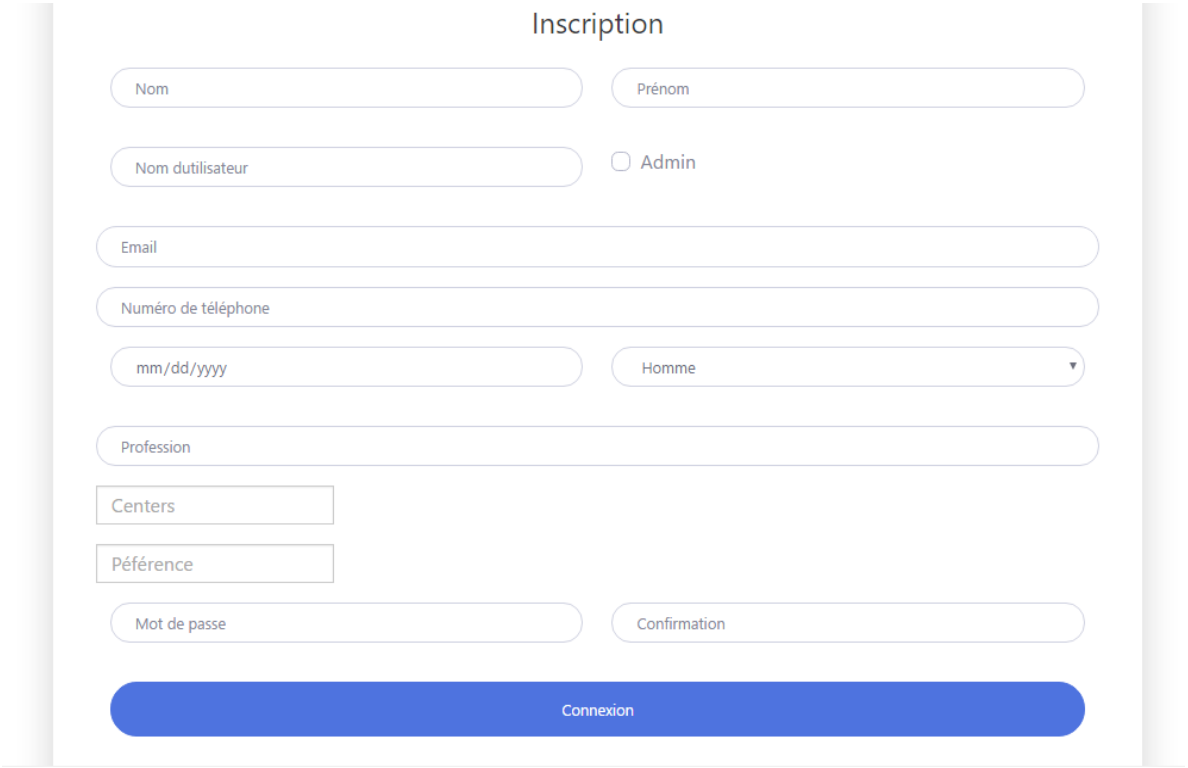

FIGURE 4.4 INSCRIPTION

- **Espace utilisateur**

Une fois identifié ou inscrit l'utilisateur accèdera à son espace personnel ou il aura accès aux différentes fonctionnalités du site.

- **Lancement des requêtes**

L'utilisateur devra exprimer son besoin d'information sous forme d'une requête dans la case adéquate.

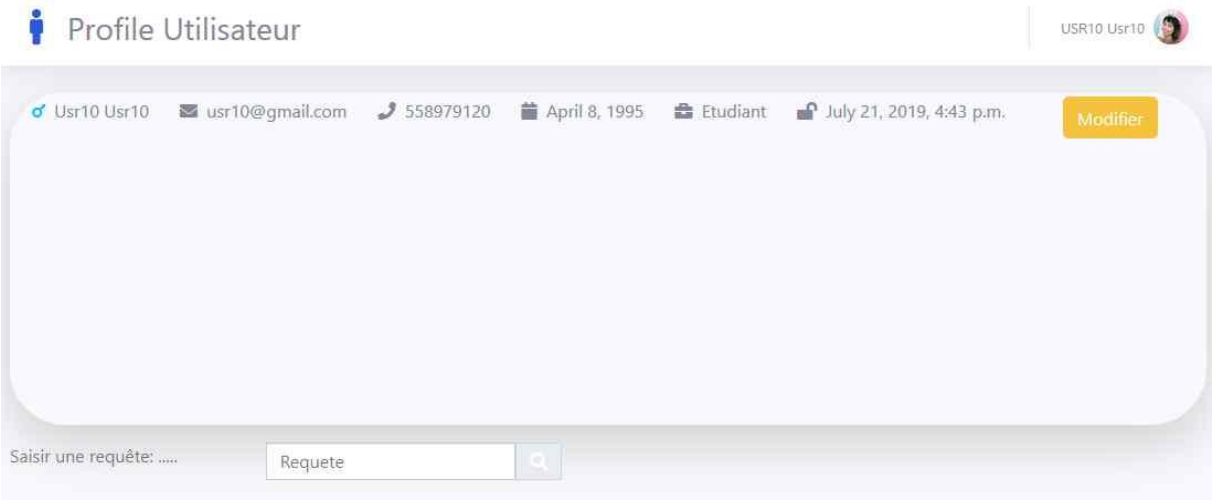

FIGURE 4.5 REQUETE

- **Consultation/Modification**

sur la barre en haut nous avons les différentes informations personnelles avec possibilité de les modifier.

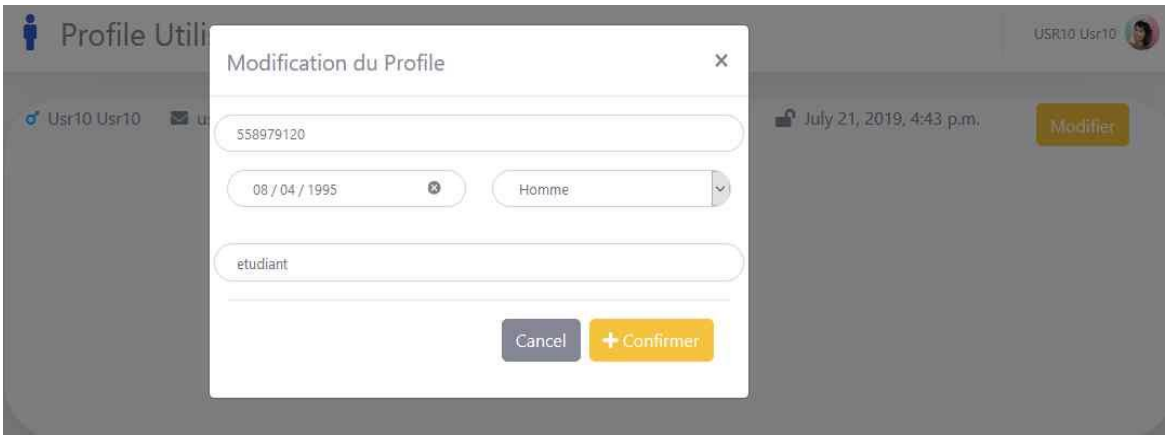

FIGURE 4.6 MODIFICATION

- **Centres d'intérêts/Préférences/Requêtes :**

Dans cette rubrique l'utilisateur pourra voir ces différents centres d'intérêts et observer ainsi leurs évolutions, pour cette partie on-a préféré les illustres sous forme d'un diagramme, pour les préférences et les requêtes ils sont affichées sous forme de tableaux accompagnés de la date et l'heure d'émission. (Possibilité de recherche d'une requête/centre/préférence précise).

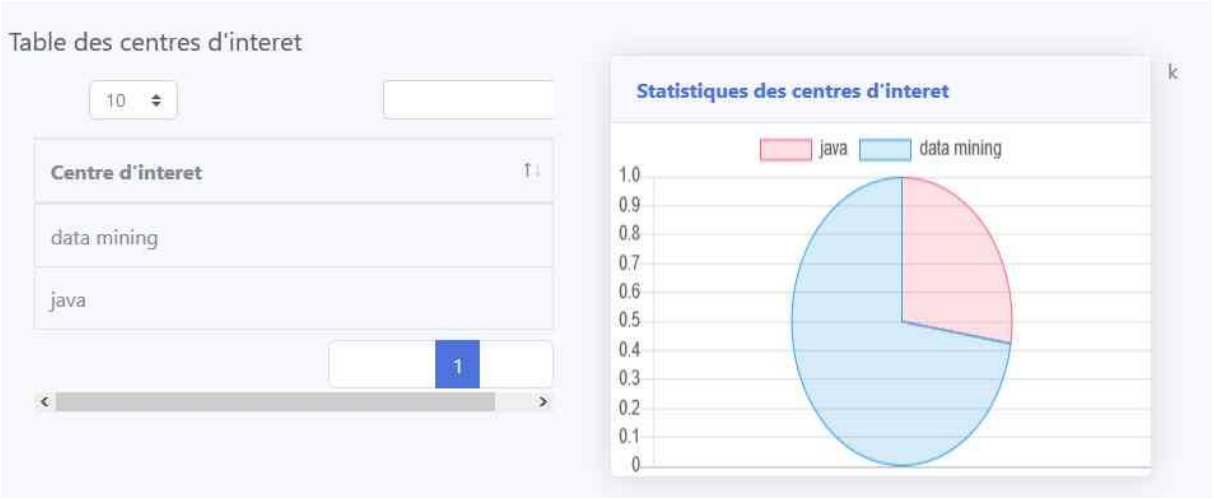

FIGURE 4.7 CENTRES D'INTERETS ET PREFERENCES

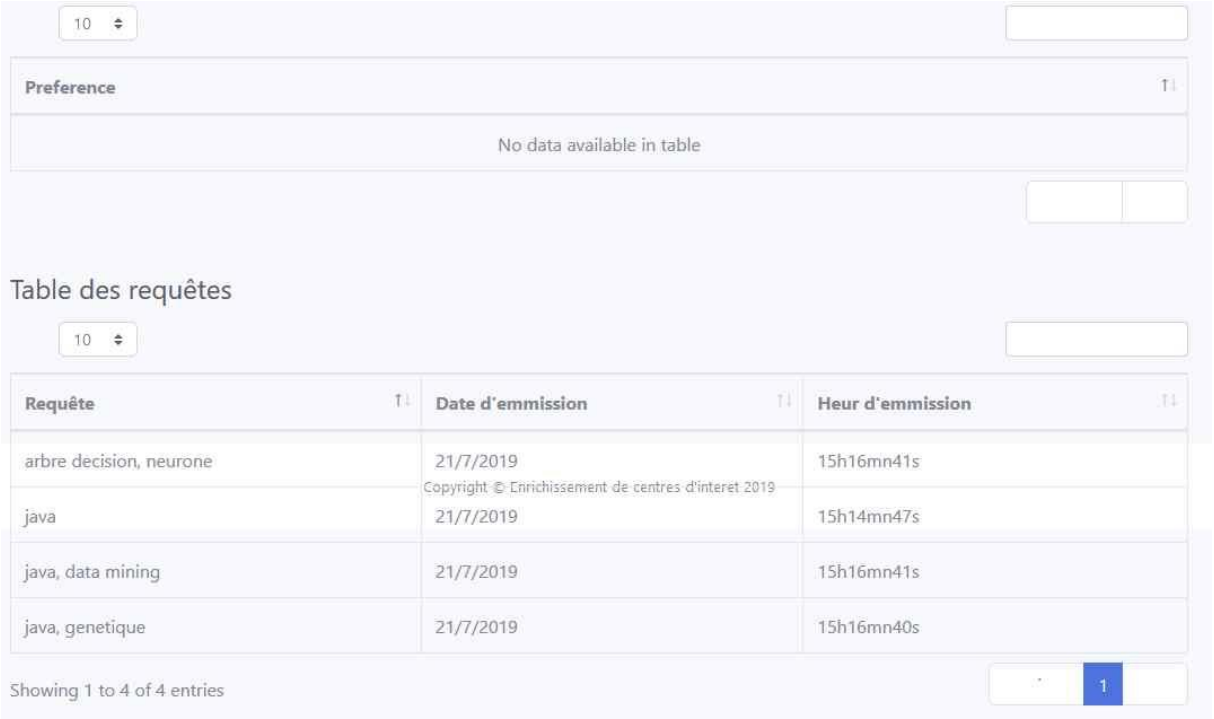

FIGURE 4.8 HISTORIQUE REQUETES

- **Administrateur**

  L'espace administrateur est réservé à la personne qui aura comme rôle la gestion du site il aura la tâche de gérer le module d'enrichissement et consulter les informations liées aux différents membres ainsi qu'un accès aux différentes statistiques.

  Lors d'accès à son espace l'administrateur verra les différentes informations liées au site d'une manière générale et il aura aussi la possibilité de :

  -Voir tous les profils inscrits sur le site (actif et non actif) et accompagné des informations générales sur ces derniers tels que son nom, prénom, dernière mise à jour et nombres de requêtes.

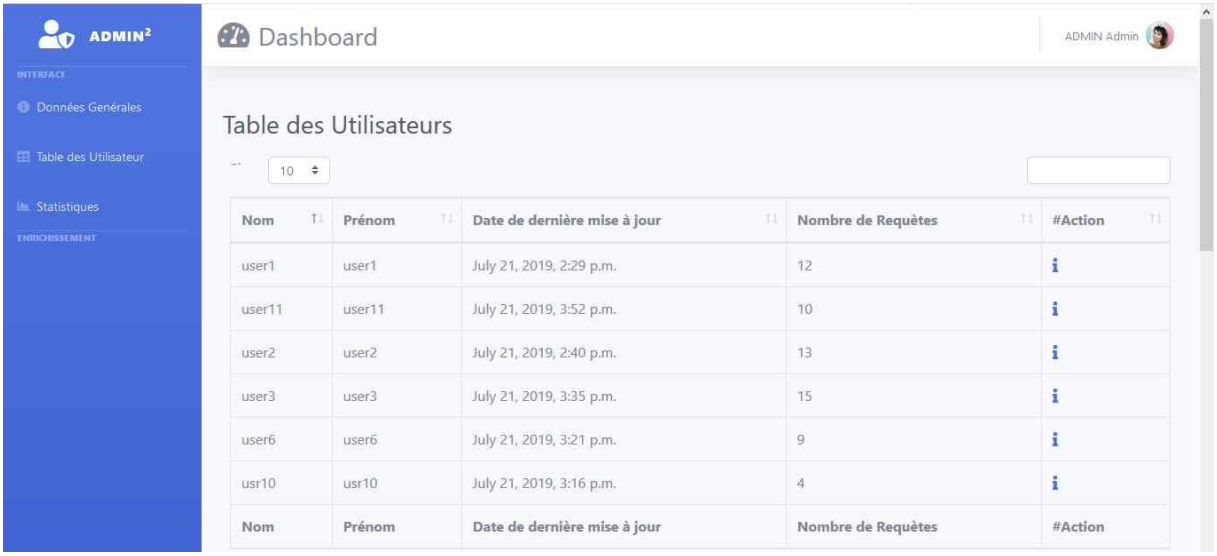

FIGURE 4.9 ESPACE ADMINISTRATEUR

Nous pourrons par la suite voir un profil spécifique en détail (en cliquant sur l'icône action), ses différentes informations personnelles, tableaux pour afficher ses centres d'intérêts/préférences (un diagramme accompagne les centres d'intérêts pour les illustrer accompagné de leur poids).

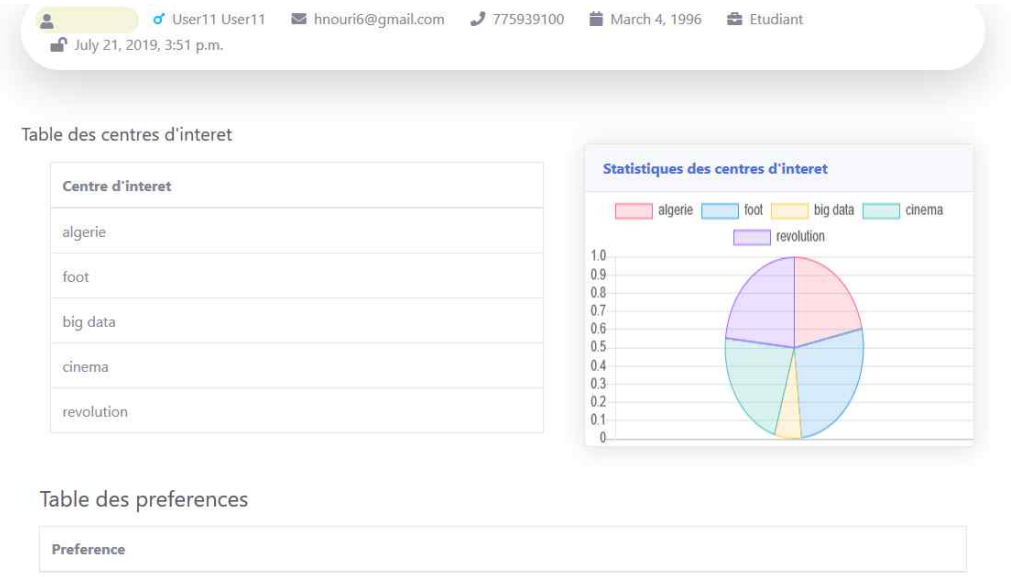

FIGURE 4.10 CONSULTATION PROFIL

Et un tableau contenant l'historique de ses requêtes.

Table des requêtes

Requête	Date d'emmission	Heur d'emmission
big data	21/7/2019	15h46mn15s
big data	21/7/2019	15h46mn15s
cinema	21/7/2019	15h46mn16s
cinema	21/7/2019	15h46mn16s
cinema	21/7/2019	15h49mn27s
foot	21/7/2019	15h49mn28s
foot	21/7/2019	15h49mn29s
revolution	21/7/2019	15h52mn18s
revolution	21/7/2019	15h52mn19s
revolution	21/7/2019	15h52mn19s

FIGURE 4.11 HISTORIQUE REQUETES PROFIL

Grace à la rubrique données l'administrateur aura une vue sur le nombre de requêtes émises sur le site ainsi que les requêtes en attente (celle qui n'ont pas été traitées pour le processus d'enrichissement depuis la dernière mise à jour)

FIGURE 4.12 DONNEES ENRICHISSEMENTS

- Enfin dans la rubrique statistiques l'administrateur aura une vue sur les centres d'intérêts recherches sur le site accompagnés d'un diagramme ainsi qu'un graphe pour illustrer les statistiques d'émission requêtes par période (quotidienne mensuelle)

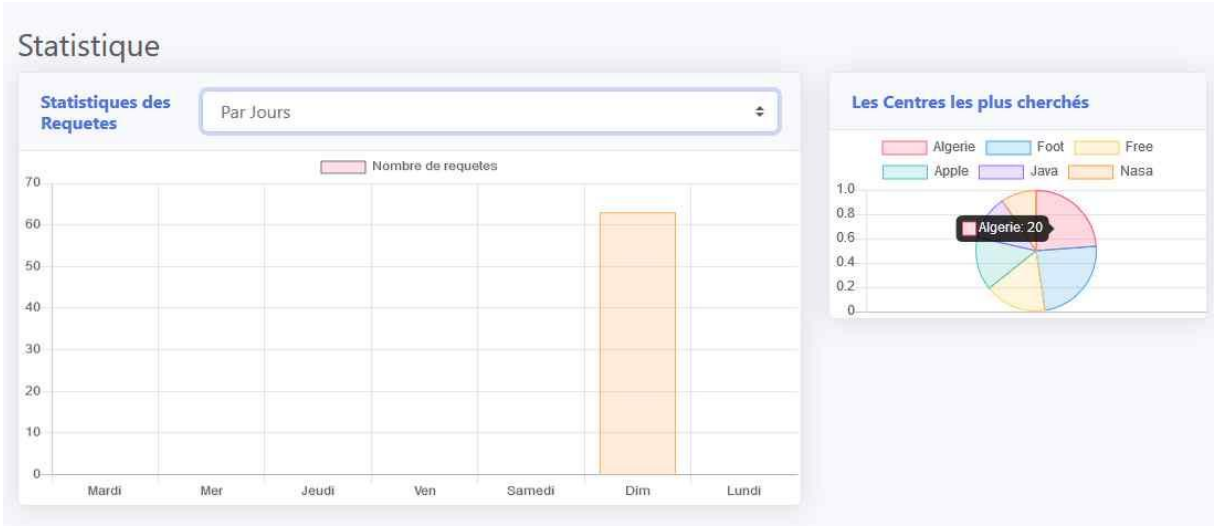

FIGURE 4.13 STATISTIQUE

- **Lancer enrichissement**

C'est le cœur de la partie administration du site pour mettre à jour les centres d'intérêt des membres en fonction de leurs requêtes. A partir du menu sur la gauche si y'a des requêtes en attente (exprimer après le processus de mise à jour) un bouton lancer enrichissement s'affichera qui déclenchera le processus de mise à jour des centres d'intérêts pour les membres actif.

FIGURE 4.14 MENU ADMIN

À la fin du processus l'administrateur pourra par membre :

-En haut visualiser toutes les requêtes émises sur un tableau.

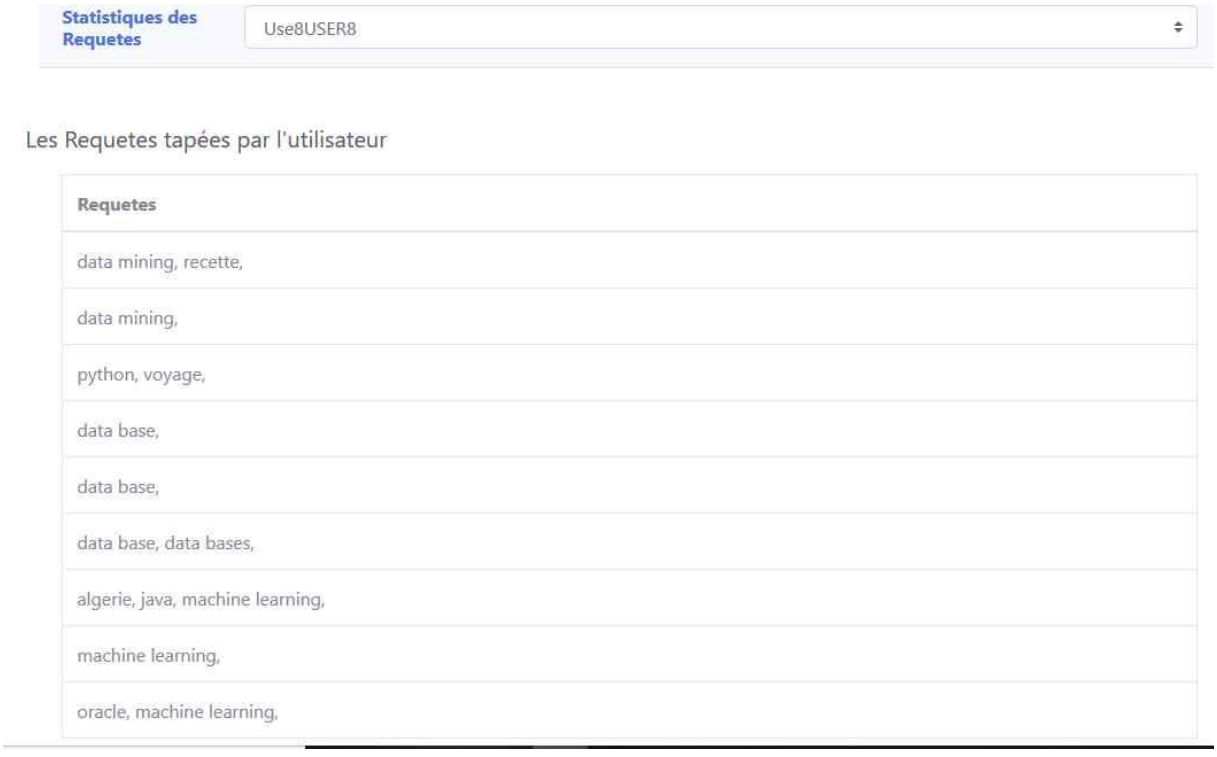

FIGURE 4.15 TABLE REQUETES

-Visualiser le profil avec les anciens centres d'intérêts et celui-ci après mise à jour accompagnes de leur poids.


Résultat de l'enrichissement

Centre d'interet avant Enrichissement	Centre d'interet après Enrichissement
algerie -> Poids = 0.1745014245014245.	algerie -> Poids = 0.21609278789247283.
foot -> Poids = 0.20192307692307693.	foot -> Poids = 0.26729888186181294.
oracle -> Poids = 0.0851392586365264.	big data -> Poids = 0.06296355703514758.
musique -> Poids = 0.0851392586365264.	cinema -> Poids = 0.2207879189392212.
data base -> Poids = 0.0851392586365264.	revolution -> Poids = 0.23285685427134553.
big data -> Poids = 0.13409433235252904.	
cinema -> Poids = 0.2340633903133903.	

FIGURE 4.16 ENRICHISSEMENT

En bas nous avons les occurrences des centres présentent dans les requêtes ainsi que l'évènement de suppression (en cas d'suppression de centres).

FIGURE 4.17 INFORMATION ENRICHISSEMENT

4. Conclusion

Au cours de ce chapitre, nous avons abordé l'implémentation de notre système et les outils utilisés pour le développement de celui-ci, ensuite nous avons entamé la présentation des différentes interfaces et fonctionnalités de notre application à travers les différentes captures, tout en détaillant les fonctionnalités les plus importantes tels que le processus d'enrichissements, consultation des utilisateurs et leur différentes statistique à travers des graphes et la phases de recherche pour les membres ainsi que la consultation du profil.

Conclusion générale

La recherche d'information est un domaine important et utile afin de fournir les différents outils dans le but de la sélection des informations pertinentes, ce domaine englobe le stockage de quantité importante d'informations et les parcourir suivant la recherche de l'utilisateur pour retourner une liste de résultats pertinentes, son principal défi est de retourner une liste de résultat qui correspond le mieux à ses attentes.

La personnalisation dans les SRI et l'introduction de la notion de profil qui est définis (par ses centres d'intérêts et préférences) et exploiter dans le processus de recherche d'information, représente un atout considérable qui servira à retourner une liste de résultats qui correspond mieux aux attentes, mais la profillisation dans les SRI nécessite le maintien du profil utilisateur à jour en suivant une multitude de facteur à travers le temps et ce pour garantir l'efficacité du système.
A cet effet, le travail réalisé à travers ce mémoire est focalisé sur l'enrichissement des centres d'intérêts de l'utilisateur suivant les requêtes exprimant son besoin d'informations exprimé par celui-ci.
Pour ce faire, nous avons opté pour une des techniques de data mining connu pour sa puissance dans la classification qui sont les réseaux de neurones artificiels, en élaborant une solution qui nous a permis d'effectuer cette mise à jour selon des critères bien précisés et en suivant une multitude de situation pour la prise de décision sur l'utilisation des centres d'intérêts des utilisateurs pour un ajout, suppression ou modification.

Quoi que nous avons réalisé le principal objectif de notre travail, mais nous envisageons quelques perspectives qui permettent l'amélioration et la conformité de notre travail notamment :

- Ajouter un réseau de neurones en plus pour la classification des requêtes au centres d'intérêts existant qui va améliorer l'efficacité de l'enrichissement
- Programmer un agent intelligent pour lancer le processus de mise à jour des membres actif en cas d'absence de l'‘administrateur
- Ajouter de nouveau cas dans le fichier d'apprentissage pour les réseaux de neurones pour l'efficacité de la classification
- Migration du site vers une base de données big-data pour permettre la mise en œuvre du site au sein d'un grand organisme

La comparaison des résultats retournés par d'autres méthodes avec à notre solution pour la phase de mise à jour

Références bibliographiques

[Abbasi, 13] M. Abbassi. "Un modèle de reformulation des requêtes pour la recherche d'information sur le Web ". université d'Ouargla. 2013

[Alaoui Ismaili, 16] O.Alaoui Ismaili, 'Clustering prédictif Décrire et Prédire simultanément', thèse de doctorat en informatique, Université Paris–Saclay-France, 2016.

[Azé, 03] J.Azé, 'Extraction des connaissances à partir des données numériques et textuelles', thèse de doctorat en informatique, Université paris-sud-France, 2003.

[Baeza & al, 99] R. Baeza-Yates et Ribeiro-Neto. «Modern Information Retrieval». New York: ACM Press; Harlow England: Addison-Wesley, cop., 1999.

[Bennani, 14] Y.BENNANI, 'Apprentissage par réseaux de neurones artificiels, support cour, École de Printemps sur l'Apprentissage artificiel', Université de Carry le Rouet-France, 2014.

[Bennett & al, 12] Bennett, P. N., R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisyuk, et X. Cui. Modeling the Impact of Short- and Long-term Behavior on Search Personalization. In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, p. 185–194. New York, NY, USA : ACM. (2012).

[Bouaziz, 17] M.Bouaziz, 'Réseaux de neurones récurrents pour la classification de séquences dans des flux audiovisuels parallèles ', thèse de doctorat en informatique, Université d'Avignon et des Pays de Vauclus-France, 2017.

[Boughanem & al, 03] M. Boughanem, d. savoy, « Recherche d'information, états des lieux et perspectives », Mermés sciences publication, 2008.

[Canut & al, 15] Canut, Marie-Françoise and On-At, Sirinya and Péninou, André and Sèdes, « Florence Enrichissement du profil utilisateur à partir de son réseau social dans un contexte dynamique : application d'une méthode

de pondération temporelle ». Institut de Recherche en Informatique de Toulouse (IRIT), Université de Toulouse. (2015)

[Challam & al, 04] V.K.R Challam. Contextual information retrieval using ontology based user profiles. In Master of science in computer science. Jawaharlal Nehru Technological University, 2004

[Chamroukhi, 12] F.Chamroukhi, 'Classification supervisée : Les K-plus proches voisins' projet-1-i41-Knn, Projet 1, Université du Sud Toulon-France, 2013-2013.

[Cheng & al, 08] Cheng, Y., G. Qiu, J. Bu, K. Liu, Y. Han, C. Wang, et C. Chen. Model Bloggers Interests Based on Forgetting Mechanism. In Proceedings of the 17th International Conference on World Wide Web, p. 1129–1130. New York, NY, USA : ACM. (2008).

[Chiaramella & al, 07] Y. Chiaramella et P. Mulhem. " De la documentation automatique à la recherche d'information en contexte". Laboratoire d'Informatique de Grenoble. 2007.

[Crabtree & al ,98] Crabtree, S. Soltysiak, M. Pp, et I. Re. (1998). Identifying and tracking changing interests. International Journal on Digital Libraries 2, 38–53.

[Daoud, 09] M. Daoud. "Accès personnalisé à l'information : approche basée sur l'utilisation d'un profil utilisateur sémantique dérivé d'une ontologie de domaines à travers l'historique des sessions de recherche" .Université Paul Sabatier - Toulouse III, 2009.

[Durand, 04] N.DURAND, 'Algorithmes génétiques et autres outils d'optimisation appliqués à la gestion du trafic aérien', thèse de doctorat en informatique et télécommunication, L'institut national polytechnique de Toulouse-France, 2004.

[El-amin, 14] R.EL-AMIN, 'Techniques de data mining pour la gestion de la relation client dans les banques', Thèse de Doctorat, Université de Mohamed khider-Biskra, Algérie, 2014.

[Ezzikouri & al, 08] H.EZZIKOURI, M.Fakir, 'Algorithmes de classification : ID3 & C4.5', support cours, Université Sultan Moulay Slimane, Maroc.

[Fayyad & al, 96] U.M. Fayyad, G. Piatesky-Shapiro, P. Smyth, and R. Uthurusamy, 'From Data Mining Advances in Knowledge Discovery', The AIII/MIT Press, 1996.

[Fiolet, 06] V.Fiolet, 'Algorithmes distribués d'extraction de connaissance', thèse de doctorat en informatique, Université science et technologie de Lille-France, 2006.

[Gauch & al, 03] S. Gauch, J. Chaffe et P. Pretschner. Ontology based user profiles for search and browsing. volume Special issue on user modelling for Web and hypermedia information retrieval, 2003

[Hadef & al, 14] M. HADEF et F. MEHAOUA « Prise en compte du profil utilisateur Dans un système de recherche d'information », Mémoire Master, Universite Kasdi Merbah-Ouargla, 2014.

[Hand, 00] D.Hand, 'Rapport sur la science et la technologie n°8, La statistique ', thèse de doctorat en informatique, Académie des Sciences Paris-France ,2000.

[Hernandez, 06] N. Hernandez. «Ontologie de domaine pour la modélisation du contexte en recherche d'information», thèse de doctorat en informatique, Université Paul Sabatier. 2006.

[Houmadi, 07] BENAMAR HOUMADI, 'Comme partielle de la maitrise en mathématiques et informatique appliquées, étude exploratoire d'outils pour le data mining'. Thèse de doctorat en informatique, Université à ROIS-RIVIÈRES-Canada, 2007.

[Kaadoud& al, 18] I.Chraibi Kaadoud & A.Garenne. Architecture des réseaux de neurones : Réseaux de neurones artificiels classiques (2/3), Publication sur le blog de http : //www.scilogs.fr/intelligence-mecanique, 2018.

[Kacem & al, 14] Kacem, A., M. Boughanem, et R. Faiz. Time-Sensitive User Profile for Optimizing Search Personlization. In V. Dimitrova, T. Kuflik, D. Chin, F. Ricci, P. Dolog, et G.-J. (2014).

[Kalakh, 13] M. Kalakh, Modélisation avec les réseaux de neurones d'un canal UWB dans un environnement minier souterrain. Mémoire. Rouyn-

Noranda, Université du Québec en Abitibi-Témiscamingue, Génie, 83 p. (2013).

[Khabzaoui, 06] M. Khabzaoui, 'Modélisation et résolution multi-objectifs des règles d'association : Application analyse de données biopyces', thèse de doctorat en informatique, Université des Sciences et Technologies de Lille-France, 2006.

[Kompaoré, 08] N. Kompaoré. "Fusion de systèmes et analyse des caractéristiques linguistiques des requêtes : vers un processus de RI adaptatif". Thèse de doctorat en informatique, Université Paul Sabatier de Toulouse, 2008

[Kostadinov, 03] Kostadinov. "La personnalisation de l'information, définition de modèle de profil utilisateur". Thèse de Master. Université de Versailles. 2003.

[Kumaran& Allan, 08] G. Kumaran, et J. Allan, « Adapting information retrieval systems to user queries, Information » Processing & Management, vol. 44, n° 6, p.1838-1862, 2008.

[Lamiche, 13] M.Lamiche, 'Fusion et fouille de données guidées par les connaissances : application à l'analyse d'image', thèse de doctorat en informatique, UNIVERSITE MOHAMED KHIDER - BISKRA, 2013.

[Lano, 2009] K. Lano, UML 2 Semantics and Applications. John Wiley & Sons, Inc., New York, NY, USA. (Cit́e dans la page 61.) (2009).

[Li & al, 13] Li, D., P. Cao, Y. Guo, et M. Lei. Time Weight Update Model Based on the Memory Principle in Collaborative Filtering. Journal of Computers 8. (2013).

[Lin & al ,05] C. Lin, G.R Xue, H.J Zeng et Y. YU. Using probabilistic latent semantic analysis for personalised web search. In Proceedings of the APWeb Conference, pages 707–711, 2005

[Maloof & al, 00] Maloof, M. A. et R. S. Michalski. Selecting Examples for Partial Memory Learning. Machine Learning 41, 27 ! 52. (2000).

[Mededjel & al, 07] M.Mededjel, H.Belbachir, 'Post-élagage Indirect des Arbres de

Décision dans le Data Mining', Conference : Sciences of Electronic Technologies of Information and Télécommunications-Tunisie, 2007.

[Mezghani & al, 15] M.Mezghani, A.Péninou, C.Amel Zayani, I.Amous, F.Sèdes. Analyse du comportment d'annotation du réseau social d'un utilisateur pour la détection des intérêts-Application sur Delicious. Revue des Sciences et Technologies de l'Information - Série ISI: Ingénieriedes Systèmes d'Information, Lavoisier, 2015, vol. 4, pp. 85-111.

[Salton & al, 83] G. Salton et McGill, M. J. W.N. "Introduction to modern information retrieval. McGraw-Hill computer science series". McGraw-Hill. 1983.

[Salton, 68] G. Salton, Automatic Information Organization and Retrieval. McGraw Hill Text, 1968.

[Santos ,15] F.Santos, 'CNRS, UMR 5199 PACEA'

[Sirinya, 17] O. Sirinya. « Temporalité et réseaux sociaux : prise en compte de l'évolution dans la construction du profil utilisateur ». Université Paul Sabatier - Toulouse III, 2017.

[Talbi, 15] E-G.Talbi, 'Fouille de données (Data Mining) -Un tour d'horizon', support cour, Laboratoire d'Informatique Fondamentale de Lille-France, 2015

[Tamine & al ,07] L. Tamine, N. Zemirli, W. Bahsoun. "Approche statistique pour la définition du profil d'un utilisateur de système de recherche d'information ». Revue I3 - Information Interaction Intelligence, Cépaduès, 2007

[Tan & al, 06] Tan, B., X. Shen, et C. Zhai. Mining Long-term Search History to Improve Search Accuracy. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p. 718–723. New York, NY, USA : ACM. (2006).

[Tir, 05] R.Tir, 'Apport de Data Mining a la Performance De L'entreprise', Conférence Paper publication sur https://www.researchgate.net/publication/290394471_APPORT_DU_DATA_MINING_A_LA_PERFORMANCE_DE_L'ENTREPRISE, 2005.

[Touzet, 92] C. Touzet, 'Les réseaux de neurones artificielles : introduction au connexionnisme', support cour, 1992.

[Wen & al, 04] J.R Wen, N. Lao etW. Y Ma. Probabilistic model for contextual retrieval. In Proceedings of the 27th annual internationsl ACMSIGIR Conference on Research and development in Information retrieval, pages 57–63, August 2004.

[Zayani, 08] C. Zayani, Contribution à la définition et à la mise en œuvre de mécanismes dadaptation de documents semi-structuré. Thèse de doctorat, Université Paul Sabatier, Toulouse, France. (2008).

[Zemirli, 08] N. Zemirli. "Modèle d'accès personnalisé à l'information basé sur les Diagrammes d'Influence intégrant un profil utilisateur évolutif". Thèse de doctorat en informatique. Université Paul Sabatier de Toulouse III. 2008.

[Zemmouri, 13] E. Zemmouri, 'Représentation et gestion des connaissances dans un processus d'Extraction de Connaissances à partir de Données multipoints de vue', thèse de doctorat en informatique, Université Moulay Ismaïl-Maroc, 2013.

[Zheng & al, 11] Zheng, N. et Q. Li. A recommender system based on tag and time information for social tagging systems. Expert Systems with Applications 38, 4575 ! 4587. (2011).